

STIC-Biotech/ChemLib

From: Li, Ruixiang
Sent: Thursday, July 11, 2002 11:37 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application NO: 09/899,513

Please do a standard search on SEQ ID NOS: 1 and 2 against the commercial nucleic acid databases.

Thank you very much!

Ruixiang Li
GAU 1646
CM1(10E18
Mail Box-10C01
306-0282

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 7/12
Date Completed: 7/15
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 22:57:41 ; Search time 1813.39 Seconds

(Without alignments)
7131.725 Million cell updates/sec

Title: US-09-899-513-1

Perfect score: 618
Sequence: 1 atggggggtcgttcgttcgc.....ggtgctcgccgacacgta 618

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 3595312

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

C	1	617.2	99.9	152947	2	AL354876	AL354876 Homo sapi
C	2	617.2	99.9	194215	2	AL607089	AL607089 Homo sapi
C	3	66	10.7	1116	6	AX068253	AX068253 Sequence
C	4	66	10.7	1119	6	AX076182	AX076182 Sequence
C	5	66	10.7	1119	6	AX138829	AX138829 Sequence
C	6	66	10.7	1119	9	HS272207	HS272207 Homo sapi
C	7	66	10.7	1119	9	AF411112	AF411112 Homo sapi
C	8	66	10.7	1325	6	AX168124	AX168124 Sequence
C	9	66	10.7	1560	6	AX352473	AX352473 Sequence
C	10	66	10.7	1720	6	AX068263	AX068263 Sequence
C	11	66	10.7	2444	6	AX305129	AX305129 Sequence
C	12	66	10.7	2480	6	AX128511	AX128511 Sequence
C	13	66	10.7	245880	2	AC079387	AC079387 Homo sapi
C	14	62	10.0	134940	2	AC018939	AC018939 Homo sapi
C	15	61.6	10.0	125020	2	AC129315	AC129315 Homo sapi
C	16	60	9.7	90003	2	AC106630	AC106630 Rattus no
C	17	58.4	9.4	1534	9	AF080214	AF080214 Homo sapi
C	18	58.4	9.4	4895	6	AR108643	AR108643 Sequence
C	19	58.4	9.4	4895	9	AF055917	AF055917 Homo sapi
C	20	58.4	9.4	4925	6	AX211769	AX211769 Sequence
C	21	58.4	9.4	11828	9	AF384819	AF384819 Homo sapi
C	22	58.4	9.4	248680	2	AC008737	AC008737 Homo sapi
C	23	58.2	9.4	78307	2	AC096500	AC096500 Rattus no
C	24	56.8	9.2	14662	1	SC663	SC663 Streptomy
C	25	56.4	9.1	39314	1	SGR300302	SGR300302 Streptomy
C	26	56.4	9.1	137007	2	AP004265	AP004265 Oryza sat
C	27	55.8	9.0	31226	1	SC6F7	SC6F7 Streptomy
C	28	55.8	9.0	155760	2	AC094490	AC094490 Rattus no
C	29	55.8	9.0	196662	2	AP001107	AP001107 Homo sapi
C	30	55.6	9.0	81673	2	AC016060	AC016060 Homo sapi
C	31	55.6	9.0	174546	2	AC087688	AC087688 Homo sapi
C	32	55.4	9.0	125020	2	AF429315	AF429315 Homo sapi
C	33	55	8.9	1842	5	AB042507	AB042507 Gallus ga
C	34	55	8.9	37445	1	SC8A6	SC8A6 Streptomy
C	35	54.8	8.9	127786	2	AC094129	AC094129 Rattus no
C	36	54.6	8.8	39896	1	SC2K31	SC2K31 Streptomy
C	37	54.6	8.8	132060	2	AC099243	AC099243 Rattus no
C	38	54.6	8.8	194177	2	AC021647	AC021647 Homo sapi
C	39	54.6	8.8	194855	2	AC093904	AC093904 Homo sapi
C	40	54.2	8.8	36816	2	AC109752	AC109752 Rattus no
C	41	54	8.7	39446	1	SC8F7	SC8F7 Streptomy
C	42	54	8.7	66597	2	AC090850	AC090850 Homo sapi
C	43	53.8	8.7	40235	1	SC210A7	SC210A7 Streptomy
C	44	53.4	8.6	85434	2	AC066610	AC066610 Homo sapi
C	45	53.4	8.6	186404	2	AC105307	AC105307 Bos tauru

ALIGNMENTS

RESULT 1
AL354876/c 152947 bp DNA linear HTG 27-JUN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-244H3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.

ACCESSION AL354876
VERSION AL354876.9 GI:14575158
KEYWORDS HTG: HTGS_PHASEL; HTGS_DRAFT.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 152947)
AUTHORS McLay, K.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Jun 28, 2001 this sequence version replaced gi:13334889.

Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>

Oy	361	acggttcacctatgatggcgcggttgcttcttcgcgcgcgcatacaagctgtgcgcgtcgaagt	420
Db	53725	ACGTTCTCCACTATAGCGGGGTGTGTCCTTCCGCCGCCTCATAGCCTGTGCCCTCGGCT	53666
Oy	421	tctaacgcgtctcccgaggccaccagcgagctgcgcgcgtgcgcgcgtgtactaagbgcccgcg	480
Db	53665	TGCTACGCGGHTCCGGGGGCCAAGGGGGCGCTGCACCCTGCCTGGCGGCTCTACGGGCCCCCG	53606
Oy	481	cgcgcgcgtcgtcttgcccttcgtccttcgctgcggcgagcttgcccccttcgccttgagaaac	540
Db	53605	CgcGcGcGcCcTtGcGcCcTtTcCcTtGcGcGcGcGcCcTtGcCcTtCcGcTtGcAgChAc	53546
Oy	541	gctgcgcaagctcgcggagcttgctgccttcgcgcacagctgtgcttcgcgcgcgccttcgcctg	600
Db	53545	GCTGGCGCAAGCTCGGGCGCTGGGCTTCGGCACAGTGATGGCTTGCGGGCCGCTTCCTGCTGG	53486
Oy	601	tgcttcgcgcgcacaagtga	618
Db	53485	TGCTTCGCGCGCCAACGTGA	53468
RESULT	2		
LOCUS	AL607089/c	194215 bp DNA linear HTG 30-JAN-2002	
DEFINITION	Homo sapiens chromosome 1 clone RP11-248I9, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.		
ACCESSION	AL607089		
KEYWORDS	AL607089, 15 GI:18476794		
SOURCE	HTG: HTGS_PHASE2; HTGS_ACTIVITERIN; HTGS_DRAFT; HTGS_FULLTOP.human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Brown, J.		
JOURNAL	Direct Submission Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 1, 2002 this sequence version replaced gi:17402315. ----- Genome Center		
COMMENT	Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquerry@sanger.ac.uk ----- Project Information Center project name: ba248I9 ----- Summary Statistics Assembly program: XCAP4; version 4.5 Sequencing vector: Plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 192819 bases at least Q40 Consensus quality: 193115 bases at least Q30 Consensus quality: 193255 bases at least Q20 Insert size: 194215; sum-of-contigs Insert size: 193590; 7.2% error; agarose-fp Quality coverage: 9.54x in Q20 bases; sum-of-contigs Quality coverage: 9.62x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 2 contigs. Gaps between the contigs * are represented as runs of N. The order of the pieces * is believed to be correct as given, however the sizes * of the gaps between them are based on estimates that have * provided by the submittor. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved. * 1 16618: contig of 16618 bp in length * 16619 16718: gap of 100 bp * 16719 194215: contig of 177497 bp in length.		

AX168124	LOCUS	AX168124	1325 bp	DNA	Linear	PAT 03-JUL-2001
AX168124	DEFINITION	Sequence 42 from Patent W00142288.				
AX168124	ACCESSION	AX168124				
AX168124.1	VERSION	GI:14597402				
AX168124.1	KEYWORDS					
AX168124.1	SOURCE	human.				
AX168124.1	ORGANISM	Homo sapiens				
AX168124.1	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AX168124.1	AUTHORS	1 (bases 1 to 1325)				
AX168124.1	TITLE	Buford,N., Baughn,M.R., Au-Young,J., Yang,J., Lu,D.A. and Reddy,R.				
AX168124.1	JOURNAL	G-protein coupled receptors Patent: WO 0142288-A 42 14-JUN-2001; Incyte Genomics, Inc. (US)				
AX168124.1	FEATURES	location/qualifiers				
AX168124.1	SOURCE	1..1325				
AX168124.1	ORGANISM	"/organism="Homo sapiens"				
AX168124.1	DB_XREF	"/db_xref="taxon:9606"				
AX168124.1	NOTE	"/note="Incyte ID No: 3168839CB1"				
AX168124.1	BASE COUNT	195 a 473 c 403 g 254 t				
AX168124.1	GCIN					
AX168124.1	Query Match	10.7%; Score 66; DB 6; Length 1325;				
AX168124.1	Best Local Similarity	50.2%; Pred. No. 0.72;				
AX168124.1	Matches 223; Conservative	1; Mismatches 207; Indels 13; Gaps 2;				
AX168124.1	OY	112 gatgacctgtcgcgttcattcgttggtccgcggcctatgctccttgagcactgagcttgagg 171				
AX168124.1	DB	90 GACTACCGACCTACCCACGCGCTGCATCTGGGTGCTTACAGCTTGtGTGCTGCGG 149				
AX168124.1	OY	172 ctgcagcaacacgttgcggcccttgccaatgttcatlccgcagcgcgcgcgccttgagccag 231				
AX168124.1	DB	150 CTCGCCCTCAACGCGCTAGCCCT--CTGGGCTTCTCGCGCGCGCTGCGCGTGCACATCG 206				
AX168124.1	OY	232 gccgcctctctactctttaaactggtcctcgtgtgattgattctttaaagtaagctcg 291				
AX168124.1	DB	207 GGGTGACGCTGATCATGTGTAACCTGGCGGCCAGCAGACTGCTTTCACCTTCGCTG 266				
AX168124.1	OY	292 cagctgtgtcctaacctactacgtg-----ggcctgcccggaggccgcctcgcaac 341				
AX168124.1	DB	267 CCCGTCGTCCTCTACTACTAGCACTGCACCACTGGCCCTTCCCGAGCCTCCTGGCGAG 326				
AX168124.1	OY	342 gcgcgcggggcccaactactaagtgctccacctatgcgcgcgcgcgcgcgcgcctcattc 401				
AX168124.1	DB	327 ACGAGGGCGCGCATTTCCAGATGAACATGTAGCGCAGCACTGCATCTCTGATGCTCAAC 386				
AX168124.1	OY	402 agcgtgtgcgcgtgagctgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 461				
AX168124.1	DB	387 AACGTGACCGCTACGCCGCCATGTGTGACCCCGCTGCGACTGCGCACCTGGCGGCGCC 446				
AX168124.1	OY	462 cgtgtcctaacgagbcccgc 521				
AX168124.1	DB	447 CGCGTGGCGCGCTCTGTGCTGGCGCGTGTGGCGCGCATCTGTGTGTTCCGTCGCC 506				
AX168124.1	OY	522 cctcctgtccttgagacacgcgtcg 545				
AX168124.1	DB	507 GCCGCCCGGTGACAGGCCCTCG 530				
AX168124.1	RESULT 9	AX352473	1560 bp	DNA	Linear	PAT 06-FEB-2002
AX168124.1	LOCUS	AX352473				
AX168124.1	DEFINITION	Sequence 25 from Patent W00190187.				
AX168124.1	ACCESSION	AX352473				
AX168124.1	VERSION	AX352473.1				
AX168124.1	KEYWORDS	GI:18617722				
AX168124.1	SOURCE	human.				
AX168124.1	ORGANISM	Homo sapiens				
AX168124.1	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AX168124.1	REFERENCE	1 (sites)				

[illegible]

* 205732 206945: contig of 1214 bp in length

Tabor, P., Taylor, T., Vasquez, R., Vinson, R., Vo, Q., Wahbah, M., Wallington, S., Weinstein, G., Weinstein, I. R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.	Direct Submission	Unpublished	2 (bases 1 to 245880)
Submitted (01-SEP-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
On Sep 1, 2000 this sequence version replaced gi:4589937 gi:4926836 gi:4589938.			
----- Genome Center -----			
Center: Baylor College of Medicine			
Center code: BCM			
Web site: http://www.hgsc.bcm.tmc.edu/			
Contact: hgsc-help@bcm.tmc.edu			
Project Information			
Center project name: J-31			
Center clone name: RP4-76J14, RP11-43J6			
----- Summary Statistics -----			
Sequencing vector: M13; L08821			
Chemistry: Dye-primer Bodyfy: 92% of reads			
Chemistry: Dye-terminator Big Dye: 7% of reads			
Assembly program: Phrap; version 0.990329			
Consensus quality: 214565 bases at least Q40			
Consensus quality: 230283 bases at least Q30			
Consensus quality: 240000 bases at least Q20			
Estimated insert size: 29807; agarose-fp estimation			
Quality coverage: 0x in Q20 bases; agarose-fp estimation			
* NOTE: Estimated insert size may differ from sequence length			
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafi_data.html)			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 42 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			
* This record will be updated with the finished sequence			
* as soon as it is available and the accession number will			
* be preserved.			
1	27068:	contig of 27068 bp in length	
*	27069	27168: gap of unknown length	
*	27169	51197: contig of 24029 bp in length	
*	51198	51297: gap of unknown length	
*	51298	89527: contig of 38230 bp in length	
*	89528	89627: gap of unknown length	
*	89628	108061: contig of 18434 bp in length	
*	108062	108161: gap of unknown length	
*	108162	129355: contig of 21194 bp in length	
*	129356	129455: gap of unknown length	
*	129456	143034: contig of 13579 bp in length	
*	143035	143134: gap of unknown length	
*	143135	160120: contig of 16986 bp in length	
*	160121	160220: gap of unknown length	
*	160221	169752: contig of 9532 bp in length	
*	169753	169852: gap of unknown length	
*	169853	177926: contig of 8074 bp in length	
*	177927	178026: gap of unknown length	
*	178027	186068: contig of 8042 bp in length	
*	186069	186168: gap of unknown length	
*	186169	192895: contig of 6727 bp in length	
*	192896	192995: gap of unknown length	
*	192996	198186: contig of 5191 bp in length	
*	198187	198286: gap of unknown length	
*	198287	203039: contig of 4753 bp in length	
*	203040	203139: gap of unknown length	
*	203140	205631: contig of 2492 bp in length	
*	205632	205731: gap of unknown length	
*	205732	206945: contig of 1214 bp in length	
*	206946	207045: gap of unknown length	

Center project name: U3982
Center clone name: 21_D_12

NOTE: This record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 877: contig of 877 bp in length
* 878 977: gap of 100 bp
* 978 1861: contig of 884 bp in length
* 1862 1961: gap of 100 bp
* 1962 2860: contig of 899 bp in length
* 2861 2960: gap of 100 bp
* 2961 3820: contig of 860 bp in length
* 3821 3920: gap of 100 bp
* 3921 4784: contig of 864 bp in length
* 4785 4884: gap of 100 bp
* 4885 5793: contig of 909 bp in length
* 5794 5893: gap of 100 bp
* 5894 6780: contig of 887 bp in length
* 6781 6880: gap of 100 bp
* 6881 7754: contig of 874 bp in length
* 7755 7854: gap of 100 bp
* 7855 8732: contig of 878 bp in length
* 8733 8832: gap of 100 bp
* 8833 9710: contig of 878 bp in length
* 9711 9810: gap of 100 bp
* 9811 10670: contig of 860 bp in length
* 10671 10770: gap of 100 bp
* 10771 11658: contig of 888 bp in length
* 11659 11758: gap of 100 bp
* 11759 12653: contig of 895 bp in length
* 12654 12753: gap of 100 bp
* 12754 13647: contig of 894 bp in length
* 13648 13747: gap of 100 bp
* 13748 14629: contig of 882 bp in length
* 14630 14729: gap of 100 bp
* 14730 15598: contig of 869 bp in length
* 15599 15698: gap of 100 bp
* 15699 16583: contig of 885 bp in length
* 16584 16683: gap of 100 bp
* 16684 17579: contig of 866 bp in length
* 17580 17679: gap of 100 bp
* 17680 18563: contig of 884 bp in length
* 18564 18663: gap of 100 bp
* 18664 19542: contig of 879 bp in length
* 19543 19642: gap of 100 bp
* 19643 20506: contig of 864 bp in length
* 20507 20606: gap of 100 bp
* 20607 21487: contig of 881 bp in length
* 21488 21587: gap of 100 bp
* 21588 22482: contig of 895 bp in length
* 22483 22582: gap of 100 bp
* 22583 23475: contig of 893 bp in length
* 23476 23575: gap of 100 bp
* 23576 24475: contig of 900 bp in length
* 24476 24575: gap of 100 bp
* 24576 25473: contig of 898 bp in length
* 25474 25573: gap of 100 bp
* 25574 26472: contig of 899 bp in length
* 26473 26572: gap of 100 bp
* 26573 27450: contig of 878 bp in length
* 27451 27550: gap of 100 bp
* 27551 28440: contig of 890 bp in length
* 28441 28540: gap of 100 bp
* 28541 29431: contig of 891 bp in length

29432 29531: gap of 100 bp
* 29532 30392: contig of 861 bp in length
* 30393 30492: gap of 100 bp
* 30493 31363: contig of 871 bp in length
* 31364 31463: gap of 100 bp
* 31464 32327: contig of 864 bp in length
* 32328 32427: gap of 100 bp
* 32428 33268: contig of 841 bp in length
* 33269 33368: gap of 100 bp
* 33369 34264: contig of 866 bp in length
* 34265 34364: gap of 100 bp
* 34365 35235: contig of 871 bp in length
* 35236 35335: gap of 100 bp
* 35336 36212: contig of 877 bp in length
* 36213 36312: gap of 100 bp
* 36313 37171: contig of 859 bp in length
* 37172 37271: gap of 100 bp
* 37272 38141: contig of 870 bp in length
* 38142 38241: gap of 100 bp
* 38242 39098: contig of 857 bp in length
* 39099 39198: gap of 100 bp
* 39199 40093: contig of 895 bp in length
* 40094 40193: gap of 100 bp
* 40194 41082: contig of 889 bp in length
* 41083 41182: gap of 100 bp
* 41183 42041: contig of 859 bp in length
* 42042 42141: gap of 100 bp
* 42142 43033: contig of 892 bp in length
* 43034 43133: gap of 100 bp
* 43134 43975: contig of 842 bp in length
* 43976 44075: gap of 100 bp
* 44076 45026: contig of 951 bp in length
* 45027 45126: gap of 100 bp
* 45127 46058: contig of 932 bp in length
* 46059 46158: gap of 100 bp
* 46159 47015: contig of 857 bp in length
* 47016 47115: gap of 100 bp
* 47116 48006: contig of 891 bp in length
* 48007 48106: gap of 100 bp
* 48307 48978: contig of 872 bp in length
* 48979 49078: gap of 100 bp
* 49079 49987: contig of 909 bp in length
* 49988 50087: gap of 100 bp
* 50088 50999: contig of 912 bp in length
* 51000 51099: gap of 100 bp
* 51100 51988: contig of 889 bp in length
* 51989 52088: gap of 100 bp
* 52089 52964: contig of 876 bp in length
* 52965 53064: gap of 100 bp
* 53065 53967: contig of 903 bp in length
* 53968 54067: gap of 100 bp
* 54068 54953: contig of 886 bp in length
* 54954 55053: gap of 100 bp
* 55054 55944: contig of 891 bp in length
* 55945 56044: gap of 100 bp
* 56045 56921: contig of 877 bp in length
* 56922 57021: gap of 100 bp
* 57022 57901: contig of 880 bp in length
* 57902 58001: gap of 100 bp
* 58002 58880: contig of 879 bp in length
* 58881 58980: gap of 100 bp
* 58981 59882: contig of 902 bp in length
* 59883 59982: gap of 100 bp
* 59983 60870: contig of 888 bp in length
* 60871 60970: gap of 100 bp
* 60971 61852: contig of 882 bp in length
* 61853 61952: gap of 100 bp
* 61953 62814: contig of 862 bp in length
* 62815 62914: gap of 100 bp
* 62915 63746: contig of 832 bp in length
* 63747 63846: gap of 100 bp
* 63847 64699: contig of 853 bp in length
* 64700 64799: gap of 100 bp

Mon Jul 15 14:32:23 2002

us-09-899-513-1.rge

Page 12

Search completed: July 13, 2002, 02:12:56
Job time: 11715 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2002, 00:58:53 ; Search time 212.92 Seconds
(without alignments)
4983.342 Million cell updates/sec

Title: US-09-899-513-1

Perfect score: 618
Sequence: 1 atggggggcgccttgcgc.....ggtcgcgcgcacacgtga 618

Scoring table: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802:*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67.6	10.9	2559	21	AAZ95039
2	66	10.7	1116	22	AAF28683
3	66	10.7	1116	24	ABA02176
4	66	10.7	1119	21	AAA64367
5	66	10.7	1119	21	AAD01119
6	66	10.7	1119	21	AAA60118
7	66	10.7	1119	22	AAD02585
8	66	10.7	1119	22	AAF66237
9	66	10.7	1119	24	AAZ98045

C	10	66	10.7	1237	24	AAZ98085	Human DNA for pote
	11	66	10.7	1325	22	AAD08837	Human G-protein co
	12	66	10.7	1560	24	AAZ19414	Human cDNA encodin
	13	66	10.7	1720	22	AAF28693	Human protein HP03
	14	66	10.7	2444	24	AAD26369	Human G-protein co
	15	66	10.7	2480	22	AAD06509	Human CON217 G pro
	16	66	10.7	3180	22	AAF25830	Human P2YL1 DNA.
	17	66	10.7	15783	22	AAZ39803	Genomic sequence #
	18	66	10.7	15783	22	AAZ90159	Human digestive sy
	19	64.4	10.4	1119	22	AAH49504	Human GTP-binding
	20	58.4	9.4	1534	21	AAH90981	Human protease-act
	21	58.4	9.4	1534	21	AAF21431	Human low adenosin
	22	58.4	9.4	1534	21	AAZ35309	Human adenosine re
	23	58.4	9.4	4895	20	AAZ32747	Human protease-act
	24	58.4	9.4	4895	21	AAZ50775	Human protease act
	25	58.4	9.4	4925	22	AAH43632	Human PAR4 coding
	26	58.4	9.4	6546	21	AAZ21433	Human low adenosin
	27	58.4	9.4	6546	21	AAZ21433	Human adenosine re
	28	54.6	8.8	114955	20	AAZ53491	Human adenosine A1
	29	54.6	8.8	114955	20	AAZ53491	Human adenosine A1
C	30	52.8	8.5	6225	20	AAZ5273	Human enzyme-relat
	31	52.8	8.5	6225	21	AAZ20843	Human multiple tar
	32	52.8	8.5	6225	21	AAA43721	Human adenosine re
	33	52.8	8.5	35384	21	AAZ21436	Human enzyme-relat
	34	52.4	8.5	117213	19	AAZ62176	HSV-2 strain SB5 C
	35	51.8	8.4	11604	22	AAD14501	Streptomyces clav
	36	51.8	8.4	15079	16	AAZ91580	S. clavuligerus cl
	37	51.8	8.4	15079	22	AAD14499	Streptomyces clau
	38	51.6	8.3	1272	20	AAZ09010	Brn-3a polynucleot
	39	51.6	8.3	1272	21	AAA29006	Human transcriptio
	40	51.4	8.3	109519	22	AAZ08693	Micromonospora DNA
C	41	51.2	8.3	15672	12	AAZ01613	Randoin receptor
	42	51.2	8.3	2529	10	AAZ93701	Sequence encoding
	43	50.4	8.2	58857	21	AAZ58471	Nucleotide sequenc
	44	50.4	8.2	65140	22	AAD17184	Streptomyces nouns
	45	50.4	8.2	125401	22	AAD17186	Streptomyces nouns

ALIGNMENTS

RESULT 1	AAZ95039	standard; cDNA; 2559 BP.
ID	AAZ95039;	
AC	AAZ95039;	
XX		
DT	15-AUG-2000 (first entry)	
XX		
DE	Human G protein coupled receptor 15334 cDNA.	
XX		
KW	G protein coupled receptor; human; signal transduction; anaemia;	
KW	neutropenia; thrombocytopenia; diagnosis; antianaemic;	
KW	immunostimulant; haemostatic; ophthalmic; antidiabetic;	
KW	cerebroprotective; neuroprotective; nootropic; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	137..1255
FT		/*tag= a
XX		
PN	WO200023588-A2.	
XX		
PD	27-APR-2000.	
XX		
PF	18-OCT-1999;	99WO-US24368.
XX		
PR	16-OCT-1998;	98US-0173869.
XX	18-OCT-1999;	99US-0173869.
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
XX		

PT Glucksmann MA, Welch NS;
XX WPI: 2000-339687/29.
DR P-PSDB: AAV79564.
XX

PT G-protein coupled receptors and the nucleic acids that encode them
XX useful for treating anemia, neutropenia and/or thrombocytopenia -
XX
PS Claim 3; Page 138-139; 140pp; English.

CC The present sequence is that of cDNA coding for the novel human
CC G protein coupled receptor (GPCR) 15334, a protein that participates
CC in signalling pathways and which is expressed in colon, pancreas,
CC tonsil, lymph node, spleen, thymus, adrenal gland, heart and
CC peripheral blood lymphocytes, including megakaryocytes and
CC erythroblasts. To obtain 15334 cDNA, an expressed sequence
CC tag (EST) was selected on the basis of homology to GPCR sequences.
CC Primers based on the EST were used to identify 15334 cDNA from a
CC spleen cDNA library. Positive clones were sequenced and the
CC overlapping fragments assembled. The invention relates to the
CC discovery of novel GPCRs, and the polynucleotides encoding them,
CC as well as to methods of using the GPCR polynucleotides and
CC polynucleotides as targets for diagnosis and treatment of
CC receptor-mediated disorders. It also relates to drug screening
CC methods using the polypeptides and polynucleotides to identify
CC agonists and antagonists for diagnosis and treatment. Methods
CC of treating a disorder by modulating the level or activity of a
CC GPCR nucleic acid, including 15334 nucleic acid, are also claimed,
CC where the disorder is selected from anaemia, neutropenia and
CC thrombocytopenia (claimed). Other conditions associated with
CC inappropriate GPCR expression and activity can also be treated,
CC e.g. retinitis pigmentosa, nephrogenic diabetes insipidus and
CC other disorders of the central and peripheral nervous systems.

CC Sequence 2559 BP; 506 A; 785 C; 648 G; 620 T; 0 other;

Query Match 10.9%; Score 67.6; DB 21; Length 2559;
Best Local Similarity 50.5%; Pred. No. 0.00023;
Matches 224; Conservative 1; Mismatches 206; Indels 13; Gaps 2;

QY 112 gatgacctgtctgcttcacgtgtgctccgcgctacgtgcttgcgtgctggtg 171
DB 185 gactacgacgtacacccgcgctgctgctgtgtgtctacagctgtgctgctgctg 244
QY 172 ctgcagcagcaagtgtgctgctgctgctgctgctgctgctgctgctgctgctg 231
DB 245 ctccctcctcaagcgtgctgctgctgctgctgctgctgctgctgctgctgctg 301
QY 232 gacctgtctctacgtgttcaacctgtgctgctgctgctgctgctgctgctgctg 291
DB 302 gtgtgtgagcgtgtacatgtgtacacctgtgctgctgctgctgctgctgctgctg 361
QY 292 cagctgtgacgtacgtacgtacgtgctgctgctgctgctgctgctgctgctgctg 341
DB 362 ccgctgtgctgtctctacgtacgtacgtacgtacgtacgtacgtacgtacgtacgt 421
QY 342 gcggcgggggacacgtacgtgttcaacctgtgctgctgctgctgctgctgctgctg 401
DB 422 acgagcggggcgctacgtacgtgtgacgtgtgacgtgtgacgtgtgacgtgtgacgt 481
QY 402 agcgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 461
DB 482 aacgtgtgacgtacgtacgtacgtacgtacgtacgtacgtacgtacgtacgtacgt 541
QY 462 cgtgtgtgacgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 521
DB 542 cgtgtgtgacgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 601
QY 522 cctccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 545
DB 602 gcgcgcgcgcgtgacagcgcctg 625

RESULT 2

AAF28683
ID- AAF28683 standard; cDNA; 1116 BP.

XX AAF28683;
XX
XX 05-APR-2001 (first entry)
XX
XX

DE Human protein HP03378 coding sequence #1.

XX Human: hydrophobic domain; immune deficiency; autoimmune disorder;
XX allergy; tissue growth; regeneration; wound healing; burn; tumour;
XX periodontal disease; thrombolytic condition; haemostatic condition;
XX infection; ss.

OS Homo sapiens.
XX
XX MO200102563-A2.
XX
XX

PD 11-JAN-2001.

PF 16-JUN-2000; 2000MO-JP03943.

PR 02-JUL-1999; 99JP-0188835.

PA (SAGA) SAGAMI CHEM. RES. CENT.
XX (PROT-) PROTEGENE INC.

PI Kato S, Kimura T;

DR WPI: 2001-071581/08.

XX P-PSDB: AAB61612.

PT New human proteins with hydrophobic domains, useful for the treatment
XX of immune disorders, tumors, allergic conditions, thrombosis and
XX microbial infection -

PS Claim 3; Pages 116-117; 153pp; English.

CC The present invention relates to human proteins (AAB61608-AAB61617) and
CC their coding sequences (AAF28679-AAF28698). The proteins of the present
CC invention have hydrophobic domains and can be used for the treatment of
CC various immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, graft-versus-host disease and
CC Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC of allergic reactions and conditions, such as asthma and in regulation of
CC haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC utility in compositions used for bone, cartilage, tendon and/or nerve
CC tissue growth or regeneration as well as wound healing and in the
CC treatment of burns. The proteins may be used in the treatment of
CC periodontal disease and in other tooth repair processes. Other uses
CC include treatment of thrombolytic and haemostatic conditions, treatment
CC or prevention of tumours and inhibiting infection by bacteria, viruses,
CC fungi and other parasites.

SO Sequence 1116 BP; 141 A; 420 C; 339 G; 216 T; 0 other;

Query Match 10.7%; Score 66; DB 22; Length 1116;
Best Local Similarity 50.2%; Pred. No. 0.00046;
Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

QY 112 gatgacctgtctgcttcacgtgtgctccgcgctacgtgcttgcgtgctggtg 171
DB 49 gactacgacgtacacccgcgctgctgctgctgctgctgctgctgctgctgctgctg 108
QY 172 ctgcagcagcaagtgtgctgctgctgctgctgctgctgctgctgctgctgctg 231
DB 109 ctccctcctcaagcgtgctgctgctgctgctgctgctgctgctgctgctgctg 165
QY 232 gacctgtctctacgtgttcaacctgtgctgctgctgctgctgctgctgctgctg 291

Query Match	10.7%	Score 66;	DB 22;	Length 1119;
Best Local Similarity	50.2%;	Pred. No. 0.0046;		
Matches 223;	Conservative 1;	Mismatches 207;	Indels 13;	Gaps 2
112 gctgacccctgcgtctctatcgltggtgcccgcgctatgctcttgatcgtggcctggg	171			
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111				
49 gactacacgcacacccacacgcgcctgacactggtggtcttaacagcttggtgctggtccgg	108			
111 111 111 111 111 111 111 111 111 111 111 111 111 111				
172 ctgcacacgaacagtgctggcgcctgagcaatgctcatccgcacgcggcgcgctggcgcaag	231			
111 111 111 111 111 111 111 111 111 111 111 111 111 111				
109 ctcgccctcaacagcgctagcctctctggtctcttcctcgcgcgctgctggctgactcg	165			
111 111 111 111 111 111 111 111 111 111 111 111 111 111				
232 gccctgctctctacacctgttcaacctgctggtgtatgagtcttcaacgctacacgctg	291			
111 111 111 111 111 111 111 111 111 111 111 111 111 111				
166 gctggtgagcggtgcatcgtgttaacctggcgacgacctgctcttaacacctctgcg	225			
111 111 111 111 111 111 111 111 111 111 111 111 111 111				
292 cagctgctgcttaacctactactgctgctgctgctgctgctgctgctgctgctgctg	341			
111 111 111 111 111 111 111 111 111 111 111 111 111 111				

Db	226	ccgcctgcctctctcctaactaagcaactgcacacactgagccctcccccagacctctgctgcag	285		
Qy	342	gcggccggggccacactactactgcttcacactatgcggcggttgctctgcgcgcgtc	401		
Db	286	acgacggcgagccatctccagatgaacatltacgagcagctcatcttcccgatgctc	345		
Qy	402	agcgctgtgcgcgtgcggtcttgtaacgcggtccggggccagggcgagctgcgcgtgc	461		
Db	346	aacgtgtagccagctacgcgcgcacatctgacacccgctgcgactgcgcacactgcggcgcc	405		
Qy	462	cgctgctacagbqcccccgccgcgcgtgcgtgcgccttcgctctgctgcgtgcggcgtgc	521		
Db	406	cgcgctgagcgagctgctgtctgcctgcgcgcgtgctggtggcgctcatctgtgttcgcgtgc	465		
Qy	522	cctccctgccttgagacagcgcgtgg	545		
Db	466	gcgcgcgcgctgacagcgcctcg	489		
RESULT 9					
ID	AAS98045	AAS98045 standard; DNA; 1119 BP.			
XX	AAS98045;				
XX	12-MAR-2002	(first entry)			
DE	Human DNA for potential G protein-coupled receptor #3.				
XX					
KW	Human; G protein-coupled receptor; GPCR; ds; GAT4; galanin receptor;				
KW	Alzheimer's disease; amyotrophic lateral sclerosis; asthma;				
KW	atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;				
KW	chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;				
KW	depression; epilepsy; macular degeneration; lymphoma; melanoma;				
KW	multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;				
KW	psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;				
KW	tuberculosis; cognition disorder; memory disorder; anorexia;				
KW	hormonal release disorder; cardiovascular activity disorder;				
KW	pain perception disorder; obesity; diabetes; obesity;				
KW	diabetes; hyperlipidaemia; stroke; gene therapy.				
OS	Homo sapiens.				
PN	WO200185791-A1.				
XX					
PD	15-NOV-2001.				
XX					
PF	11-MAY-2001; 2001WO-US15332.				
XX					
PR	11-MAY-2000; 2000US-203217P.				
XX	18-MAY-2000; 2000US-205945P.				
PA	(LIFE-) LIFESPAN BIOSCIENCES INC.				
PI	Brown JP, Miller M, Burner G, Fabre-Suver C, Pritchard D;				
DR	WPI; 2002-066595/09.				
XX					
PT	Novel G protein-coupled receptor polypeptides including galanin				
PT	receptor polypeptides useful for identifying modulators that are useful				
PT	for treating Alzheimer's disease, psoriasis, melanoma, multiple				
XX	sclerosis, stroke				
PS	Disclosure; Page 113-114; 144pp; English.				
CC	The invention relates to an isolated polypeptide encoded by a				
CC	nucleic acid molecule that is at least 80% identical to the G				
CC	protein-coupled (GPCR) polynucleotides included in the specification.				
CC	Also included are probes based on the GPCR sequences (including				
CC	antisense probes), a host cell comprising an expression vector comprising				
CC	the GPCR sequence, antibodies raised against the polypeptides,				
CC	and methods of identifying modulators of the polypeptides. The				
CC	polypeptides are useful for identifying modulator compounds which				


```
Db 1086 GACTACGACCTACCTACCAACGCGCTGACTGTGGTGTACACCTTGCTGGCGCGG 1027
Oy 172 ctgcagcaacgacgtgagcgccctggaatgtcatccgagcgagcgccctggccag 231
Db 1026 CTCCCTCAACGCGCTAGGCTCT--CTGGCTCTTCTGCGCGCTGCGTGCACACG 970
Oy 232 gccctgtctctcctccttcaacctgtacgtctgtgtatgagttcttcacgtacgtg 291
Db 969 GTGGTAGCGCTGTACATGTACCTGGCGCCAGCGACCTGCTCTTCACCTCTCGGTG 910
Oy 292 cagctgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 341
Db 909 CCGCTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 850
Oy 342 ggcgcggcgagcaccctcctcctcctcctcctcctcctcctcctcctcctcctc 401
Db 849 ACGACGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 790
Oy 402 agcgtgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 461
Db 789 AACGTGACCCCTTACCGCGCCATGCTGACCCGCTGCGACTGCGCCACCTCGCGGCC 730
Oy 462 cgtgtcctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 521
Db 729 CGCGTGGCGCGGCTGCTCTGCTGCGCGCTGCGCGCTGCTGCTGCTGCTGCTG 670
Oy 522 cctcctcgtcgtgagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 545
Db 669 GCGCGCGCGCTGACAGCGCCCTCG 646

RESULT 11
AAD08837
ID AAD08837 standard; cDNA; 1325 BP.
XX
AC AAD08837;
XX
DT 04-SEP-2001 (first entry)
XX
DE Human G-protein coupled receptor-3 (GCR3-3) cDNA.
XX
KW Human; G-protein coupled receptor-3; GCR3-3; gene therapy; cirrhosis;
KW transgenic animal; proliferative disorder; actinic keratosis; hepatitis
KW leukemia; adenocarcinoma; lymphoma; melanoma; myeloma; epilepsy; stroke;
KW neurological disorder; Alzheimer's disease; Parkinson's disease; nausea;
KW Huntington's disease; multiple sclerosis; dementia; angina pectoris;
KW central nervous system disorder; cardiovascular disorder; hypertension;
KW atherosclerosis; congestive heart failure; gastrointestinal disorder;
KW dysphagia; peptic esophagitis; spasm; gastritis; anorexia; pyrosis;
KW pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;
KW inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;
KW Addison's disease; allergy; asthma; diabetes mellitus; antithyroid;
KW atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;
KW rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;
KW metabolic disorder; obesity; nootropic; prozoacide; virucide; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 42..1310
FT /tag= a
FT /product= "Human GCR3-3 protein"
XX
XX WO200142288-A2.
XX
XX 14-JUN-2001.
XX
XX 07-DEC-2000; 2000WO-US33382.
XX
XX 10-DEC-1999; 99US-0172852.
XX
XX 22-DEC-1999; 99US-0171732.
```

```
PR 14-JAN-2000; 2000US-0176148.
PR 21-JAN-2000; 2000US-0177331.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Burford N, Baughn MR, Au-Young J, Yang J, Lu DM, Reddy R;
XX WPI: 2001-381635/40.
XX P-PSDB: AAE04547.
XX
XX New human G-protein coupled receptor polypeptides for diagnosing,
XX preventing, and treating cell proliferative, neurological,
XX cardiovascular, gastrointestinal, autoimmune and metabolic disorders -
XX Claim 5; Page 160-161; 175pp; English.
XX
XX The present sequence is human G-protein coupled receptor-3 (GCR3-3)
XX cDNA. GCR3 is useful in somatic or germline gene therapy to correct a
XX genetic deficiency, to express a conditionally lethal gene product and
XX to express a protein which affords protection against intracellular
XX parasites and also for diagnosis of disorders associated with expression
XX of GCR3. GCR3 is also useful for generating hybridisation probes useful
XX in mapping the naturally occurring genomic sequences and to create
XX knockin humanised animals (pigs) or transgenic animals (mice or rats) to
XX model human diseases. GCR3 is used to diagnose, prevent and treat
XX proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,
XX hepatitis and cancer); cancer (breast, bladder, bone marrow, brain,
XX uterus cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma)
XX neurological disorders (epilepsy, stroke, Alzheimer's, Huntington's,
XX Parkinson's disease, multiple sclerosis, dementia and other central
XX nervous system disorders); cardiovascular disorders (angina pectoris,
XX hypertension, atherosclerosis, congestive heart failure);
XX gastrointestinal disorders (dysphagia, peptic esophagitis, oesophageal
XX spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,
XX pyrosis, pancreatitis, Crohn's disease, diarrhoea); autoimmune/
XX inflammatory disorders (acquired immunodeficiency syndrome (AIDS),
XX Addison's disease, allergies, anaemia, asthma, diabetes mellitus, atopic
XX dermatitis, glomerulonephritis, Grave's disease, osteoarthritis,
XX psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,
XX parasitic, protozoal and helminthic infections) and metabolic disorders
XX (obesity, osteoporosis, viral infections).
XX
XX Sequence 1325 BP; 195 A; 473 C; 403 G; 254 T; 0 other;
XX
XX
XX Query Match 10.7%; Score 66; DB 22; Length 1325;
XX Best Local Similarity 50.2%; Pred. NO. 0.00046;
XX Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;
Oy 112 gatgacccctcgtcgttctatcgtgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgc 171
Db 90 gactacgacacacacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 149
Oy 172 ctgcagcaacgacgtgagcgccctggaatgtcatccgagcgagcgccctggccag 231
Db 150 ctcccccacacagcgtagcct--ctgggtcttcctcgcgcgcgcgcgcgcgcgcgc 206
Oy 232 gccctgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 291
Db 207 gtgtgagcgtgtacatgtgtacacgtgcgcgcgcgcgcgcgcgcgcgcgcgcgc 266
Oy 292 cagcgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 341
Db 267 ccggtgtctctcctcctcctcctcctcctcctcctcctcctcctcctcctcgcag 326
Oy 342 ggcgcggcgagcaccctcctcctcctcctcctcctcctcctcctcctcctcctc 401
Db 327 acgagcgcgacatcttcagatgacatgtacagcagcgtgatcttcctatgtcctc 386
Oy 402 agcgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 461
Db 387 aacgtgacgcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 446
```


polypeptides and polynucleotides. GPCR polypeptides are useful for screening compounds that modulate their activity. They are useful in the diagnosis, prevention and treatment of disorders which include cell proliferative disorders such as arteriosclerosis, hepatitis, myelofibrosis, psoriasis and cancer including adenocarcinoma, leukemia, lymphoma; neurological disorders such as epilepsy, ischaemic cerebrovascular disease, Alzheimer's disease, Pick's disease, dementia, Parkinson's disease, ataxias, multiple sclerosis, bacterial and viral meningitis, Creutzfeldt-Jakob disease, schizophrenic disorders, amnesia; cardiovascular disorders such as arteriovenous fistula, atherosclerosis, hypertension, vascular tumours, myocardial infarction, hypercardiac heart disease, infective endocarditis, cardiomyopathy, myocarditis; gastrointestinal disorders such as dysphagia, peptic oesophagitis, emesis, anorexia, nausea, peptic ulcer, cholelithiasis, diarrhoea, constipation, acquired immune deficiency syndrome (AIDS), hepatic encephalopathy; autoimmune/inflammatory disorders such as Addison's disease, allergies, spondylitis, amyloidosis, anaemia, asthma, contact dermatitis, Crohn's disease, diabetes mellitus, Goodpasture's syndrome, emphysema, Grave's disease, gout, multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus, uveitis, viral, bacterial, fungal, parasitic, protozoal and helminthic infections and trauma; metabolic disorders such as diabetes, obesity and osteoporosis; and viral infections such as infection caused by viral agent classified as adenovirus, arenavirus, bunyavirus. Polynucleotides of the invention are useful as probes for assessing toxicity of test compounds. They are also used in gene therapy. The present sequence is human G-protein coupled receptor 1 (GCRBC-1) cDNA.

Sequence 2444 BP; 446 A; 803 C; 640 G; 555 T; 0 other;

Query Match 10.7%; Score 66; DB 24; Length 2444;
Best Local Similarity 50.2%; Pred. No. 0.00047;

Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

```

OY 112 gatgaccctgtcgtctcattcgtggtccgcggtcctatgcttgccacggcgccctggg 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 398 gactacagcactaccacgcgcgtcgtggtgtctacagctgtgtgctggtccggg 457
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 172 ctgacagcaacgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 458 ctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 232 gccctgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 291
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 515 gtgtgtgagcgtgtacatgtgtacatgtgtacatgtgtacatgtgtacatgtgtacatgt 574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 292 cagctgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 575 cccgtcgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 634
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 342 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 635 acgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 694
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 402 agcgtgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 695 aacgtgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 754
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 462 cgtgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 755 cgcgtgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 814
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 522 cctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 545
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 815 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 838
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 15
AAD06509
ID AAD06509 standard; DNA; 2480 BP.
XX
AC AAD06509;

```

XX 10-AUG-2001 (first entry)
DT
XX
XX Human CON217 G protein-coupled receptor DNA.
DE
XX
XX Human: G protein-coupled receptor; GPCR; CON217 protein; schizophrenia;
KW neuroleptic; nootropic; neuroprotective; bipolar disease; psychotropic;
KW neurological disorder; psychiatric disease; neurosis; anxiety; neuritis;
KW attention deficit hyperactivity disorder; neurasthenia; senile dementia;
KW affective disorder; neuropathy; Alzheimer's disease; Parkinson's disease;
XX depression; migraine; genetic screening; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 42..1160
FT /tag="a
FT /product="Human G protein-coupled receptor protein"
FT
XX
XX MO200131014-A2.
XX
XX 03-MAY-2001.
XX
XX 27-OCT-2000; 2000MO-US29601.
XX
XX 27-OCT-1999; 99US-0427653.
XX 27-OCT-1999; 99US-0427859.
XX 27-OCT-1999; 99US-0428020.
XX 27-OCT-1999; 99US-0428114.
XX 28-OCT-1999; 99US-0429517.
XX 28-OCT-1999; 99US-0429555.
XX 28-OCT-1999; 99US-0429676.
XX 28-OCT-1999; 99US-0429695.
XX 03-DEC-1999; 99US-0454359.
XX 12-JAN-2000; 2000US-0481794.
XX
XX (PHMA ) PHARMACIA & UPJOHN CO.
XX
XX Vogel G, Wood LS, Merchant K;
XX
XX MPI; 2001-328653/34.
XX P-PDB; AAE02500.
XX
XX Seven transmembrane receptor polypeptides and polynucleotides, useful
PT for treating neurological or psychiatric disorders, e.g. schizophrenia,
PT as well as for identifying compounds useful for treating schizophrenia
PT
XX
XX Claim 22: Page 201-203; 215pp; English.
XX
XX The invention relates to human G protein-coupled receptor (GPCR) and
XX their corresponding DNA molecules. GPCR is also referred as seven
XX transmembrane receptor. G protein-coupled receptor protein is useful for
XX treating neurological disorder, particularly schizophrenia. GPCR protein
XX is also useful for identifying compounds useful for treating
XX schizophrenia. These compounds are also useful for treating other
XX neurological and psychiatric diseases, e.g. depression, anxiety, bipolar
XX disease, affective disorders, attention deficit hyperactivity disorder/
XX attention deficit disorder, epilepsy, neuritis, neurasthenia, neuropathy,
XX neurosis, Alzheimer's disease, Parkinson's disease, migraine and senile
XX dementia. The invention also provides genetic screening procedures that
XX entail analysing a person's genome with respect to GPCR. The vectors are
XX useful for the recombinant production of the GPCR's. The present DNA
XX sequence encodes human CON217 G protein-coupled receptor (GPCR) protein.
XX
XX Sequence 2480 BP; 507 A; 759 C; 622 G; 592 T; 0 other;

```

Query Match 10.7%; Score 66; DB 22; Length 2480;
Best Local Similarity 50.2%; Pred. No. 0.00047;
Matches 223; Conservative 1; Mismatches 207; Indels 13; Gaps 2;

OY 112 gatgaccctgtcgtctcattcgtggtccgcggtcctatgcttgccacggcgccctggg 171

Db	90	gactacgactactaaccaacgcgcgtgcacttggctgtccacagccttggctgtgcgcgcgg	149
Qy	172	ctgccaagcaacatgtagcgagccctgtagcaattcattccagcagcgagcgccctggagccag	231
Db	150	ctccccctcaacgagcgctagccct---ctgggtctctctctgtagcgagcgagcgctgacatcg	206
Qy	232	gcccctgctctctctacacgtttccaaacgtgcgtctctgtgtgaatgagttctctacagtcacgtg	291
Db	207	gtgggtgaagcgtgtacacagttgaacccttggcgcagcagacctgtccttcaaccctctccgtg	266
Qy	292	cagctgtgagctcaactactactctg-----gagcttggcccgagagcgccgcctggcac	341
Db	267	cccgctctgcctctctctactactacgacatgcaaccacatgagcctctcccgagcctctctgtgcag	326
Qy	342	gagcgagcgagccacactactaagttatccacactatgtagcggtgtgctctggcgcgcgcacatc	401
Db	327	acgaaagcgagcgcacactctccagatgaataatgtacgacgacgcatctctccgatgcatc	386
	402	agcgctgtgacgcgtgcgcgcttcgtatcagcgcgtctcccgagcccaagggcgagcgtccgcgcctgagc	461
Db	387	aacgtgagacgcctlaagcgcgcacatcgttgcaaccgcgtgtgacgtgcgcacacgtgvcgcgc	446
Qy	482	cgggtgacctaacgbbgccccgcgcgcgcgcgcgccttgcgccttcgcgcgtgcgcgcgcgcgcgc	521
Db	447	cggctgtgagcgcgcgtcgtctgctctggagcgtgtgtagcgtctacatcctgtgttttgcgcgtgcc	506
Qy	522	ctccctctgcctggaggaacacgctg	545
Db	507	gcccgcgcgcgtgtgacaaagccttcg	530

Search completed: July 13, 2002, 02:11:42
Job time: 4369 sec

REFERENCE 1 (bases 1 to 466)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 978 Std Error: 0.00
 Seq primer: -400p from Gibco
 High quality sequence stop: 431.
 Location/Qualifiers

FEATURES

1..466
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1874835"
 /clone_1lb="NCI-CGAP_C08"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 89 a 153 c 177 g 45 t 2 others
 ORIGIN

Query Match 74.2% Score 458.6; DB 9; Length 466;
 Best Local Similarity 99.1% Pred. No. 1.9e-69;
 Matches 459; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

98 gctgaagccttgatgacccctgctgctcatcgtgtgctccgagcctatgcttgg 157
 466 gctgaagccttgatgacccctgctgctcatcgtgtgctccgagcctatgcttgg 407
 158 cacttgagccttgatgacccctgctgctcatcgtgtgctccgagcctatgcttgg 217
 406 cacttgagccttgatgacccctgctgctcatcgtgtgctccgagcctatgcttgg 347
 218 gggcctcctgagcagcctgcttctctctacgtgttcaacctggtctgttgatgtct 277
 346 gggcctcctgagcagcctgcttctctctacgtgttcaacctggtctgttgatgtct 287
 278 tcacgactgaagctgaagctgtgctgaagctgaagctgtgctgaagctgtgctgaag 337
 286 tcacgactgaagctgaagctgtgctgaagctgtgctgaagctgtgctgaagctgtg 227
 338 ccacgagcctgagcagcctgcttctctctacgtgttcaacctggtctgttgatgtct 397
 226 ccacgagcctgagcagcctgcttctctctacgtgttcaacctggtctgttgatgtct 167
 398 catcagcgtgtgctgagcctgcttctctctacgtgttcaacctggtctgttgatgtct 457
 166 catcagcgtgtgctgagcctgcttctctctacgtgttcaacctggtctgttgatgtct 107
 458 ggcctcctcctgagcagcctgctgagcctgctgagcctgctgagcctgctgagcct 517
 106 ggcctcctcctgagcagcctgctgagcctgctgagcctgctgagcctgctgagcct 47
 518 ggcctcctcctgagcagcctgctgagcctgctgagcctgctgagcctgctgagcct 560

Db 46 GGGCCCTCCCTGCTGAGACACCGCTGGCAGCTCGGGCTG 4

RESULT 4
 LOCUS A1264420/c
 DEFINITION A1264420 460 bp mRNA linear EST 28-JAN-1999
 qk54f04.x1 NCI CGAP_C08 Homo sapiens CDNA clone IMAGE:1872799 3'
 similar to SM:P2Y4_HUMAN P51582 URIDINE NUCLEOTIDE RECEPTOR ;, mRNA
 sequence.
 ACCESSION A1264420
 VERSION A1264420.1 GI:3872623
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 460)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1000 Std Error: 0.00
 Seq primer: -400p from Gibco
 High quality sequence stop: 449.
 Location/Qualifiers

FEATURES

1..460
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1872799"
 /clone_1lb="NCI-CGAP_C08"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 87 a 151 c 179 g 42 t 1 others
 ORIGIN

Query Match 73.9% Score 456.6; DB 9; Length 460;
 Best Local Similarity 99.3% Pred. No. 4.1e-69;
 Matches 457; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

112 gatgaacctgctgctcatcgtgtgctccgagcctatgcttggcctgagcctgagc 171
 460 gatgaacctgctgctcatcgtgtgctccgagcctatgcttggcctgagcctgagc 401
 172 ctgcaagcctgagcagcctgcttctctctacgtgttcaacctggtctgttgatgtct 231
 400 ctgcaagcctgagcagcctgcttctctctacgtgttcaacctggtctgttgatgtct 341
 232 gccctcctcctgagcagcctgctgagcctgctgagcctgctgagcctgctgagcct 291
 340 gccctcctcctgagcagcctgctgagcctgctgagcctgctgagcctgctgagcct 281
 292 cagctgtgctcactactactgagcctgctgagcctgctgagcctgctgagcctgctg 351
 280 cagctgtgctcactactactgagcctgctgagcctgctgagcctgctgagcctgctg 221

OY	352	ccactctactagatgltcacacccatgcaggcggtgtgtgtccttcgccggcgtcatcaagggtgtgccc	411
Dd	220	CCACTTACTAGGTGTCACCTATGAGGGGGGtghfTTTCGCCGCCTCATCAAGCTGTGCC	161
OY	412	gctgcgctctgtctacgcggtgtccggggccccaaaggcgagtctgcgcgcctgtggcccgtgtacctac	471
Dd	160	GCTGCGCTTGTAAGCGCGGTCCCGGGGCCAAGGCGGGCTGCCGCCTGTGGCCGCGTGTACC	101
OY	472	ggbgaccggcgcgagctgtcctgtgcgccttcgccttgctgtgtagcgagctggccctccctggcc	531
Dd	100	GGCGCCCCGCGCGCCTCTCCTCGCGCTTCGCTGGCTGTGGCGCGCTGTGGCCCTTCCTGCCC	41
OY	532	tggagcacgcgctgtgacaagctcggcggtgcctgcctgcgcgcac	571
Dd	40	TGAACACCGCTGGGCAAGCTCGGGGCTGTGGCTCGCCAC	1
RESULT	5	Bf000091	449 bp mRNA linear EST 06-OCT-2000
GENE		7h17dl.x1 NCI CGAP Co16 Homo sapiens CDNA clone IMAGE:3316245	3'
DEFINITION		similar to contans PTR5.t3 TAR1 repetitive element ; , mRNA sequence.	
ACCESSION		BF000091	
VERSION		BF000091.1	GI:10700366
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Euthera; Primates; Carnivora; Homnidae; Homo.	
TITLE		I (bases 1 to 449)	
JOURNAL		NCI-CGAP htp://www.ncbi.nlm.nih.gov/ncicgap.	
COMMENT		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)	
		Contact: Robert Strausberg Ph.D.	
		Email: cgaaps-remai.nih.gov	
		Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.	
		cDNA Library Preparation: M. Bento Soares, Ph.D.	
		cDNA Library Arrayed by: Greg Lennon, Ph.D.	
		DNA Sequencing by: Washington University Genome Sequencing Center	
		Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB, send email to: info@lmlb.lnl.gov	
		Seq primer: -400P from Glibco High quality sequence stop: 388.	
FEATURES		Location/Qualifiers	
source		1..449	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone_image="3316245"	
		/clone_lib="NCI_CGAP_Co16"	
		/tissue_type="Colon tumor, RER"	
		/lab_host="DH10B"	
		/note="Organ: colon; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not i; Site.2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Co16 was prepared, and ss circles were made in vitro. Following RAP hybridization, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1057416-1061255, and 1144584-114531).	
		Subtraction by Bento Soares and M. Fatima Bonaldo.	
BASE COUNT		87 a 150 c 172 g 40 t	
ORIGIN			

Query Match	72.38;	Score 446.6;	DB 10;	Length 449;
Best Local Similarity	99.68;	Pred. No. 2,1e-67;		
Matches 447;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;

OY	98	gctgaagcctctggatgaaaccccttcgtcttcatgcatacgtagcccgaggactgacctgg	157
Db	449	GCTGAAGCCTCTGGATGAAACCATGCTCGCTTCATGTGATGCCCGGACTTATGCCTTTGG	390
OY	158	cacttgagccttgaggagtgcacaagcaacgtagcgaccctggcaattcatccgacagcgcg	217
Db	389	CACTTGAGCCTTGAGGAGTGCACCAACGTAAGGCGCCCTGCAATGTTCATCCGACAGCGGCG	330
OY	218	ggagcctgggccaaagccctctctcttaectgtttcaaacctggtcctgttgtatgagtctc	277
Db	329	GGCGCCTGGGCGCACAGCCCTCGCTTCCTCAACCGTTCAACACTGGCTGCTGGTAGATGATTCT	270
OY	278	tcaagctcaagcttgagaagcttggtgctcacactiaactcctggagcgtggccggagcgccctg	337
Db	269	TCAAGCTCAAGCTTGAGAAGCTTGGTGTGGCTCAACTCACTAAGCTBGGGCTTGCCCGAGAGCGCCCTG	210
OY	338	ccaagcggcgccggggcgacactactaagtgltccaactatgcyggcgglygtcttcgcgcgct	397
Db	209	CCAGCGGGCGGGGGCGCACCTACTAAGTGTCACACCTATCGCGCGGNGHCTTCGCGCGGCGCT	150
OY	398	catcagcgtgtgcgcgctgcgcttcgtatcgcggttcggcccgagggaggtgtgcgcgct	457
Db	149	CATCAGCGGTGTGCCCTCGGCTTCGTCAAGCGGTCGCCGCGCCAGAGGCGCTGCCCGCGCT	90
OY	458	ggccccggtagcctaagbgccccggcgcgcgctgcgctgcgcttcgccttcgcttcgacggcgggcct	517
Db	89	GGCCCCGGTGTCTACAGGGCGCCCGCGCGCGCGCTTCGCGCTTCGCTGGCTGGCGGGGCGCT	30
OY	518	ggcccctccctcctgctcgtgagcaacgctggg	546
Db	29	GGCCCCCTCCCTGCTGTGAGACACCGCTGGG	1
RESULT	6		
LOCUS	Bf108800/C	392 bp	mRNA linear EST 20-OCT-2000
DEFINITION	715h02.x1 Soares NSF.F8.W OT_PA.P-S1 Homo sapiens CDNA clone IMAGE:3525338 3', mRNA sequence.		
ACCESSION	Bf108800		
VERSION	Bf108800.1	GI:10938490	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 392)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strusberg, Ph.D. Email: cgsbds-tr@mail.nih.gov This clone is available royally-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40UP from Glibo High quality sequence stop: 358.		
FEATURES	Location/Qualifiers		
source	1..392		

labeled DNA fragments were ligated into a modified polylinker, Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBSRP pool 1: 109384-310919, 333208-325855 Soares NBSRP pool 1:

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Oligo-dt track not found. Not I site shown in beginning of sequence
 is likely internal to the message. CDNA library preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 Information can be found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bbrp/image/image.html The following repetitive
 elements were found in this cDNA sequence: 236-273,
 >GC-rich#low-complexity
 Seq primer: M13 Forward
 POLYA-NO.

FEATURES

source

1. .333
 location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2709445"
 /clone_lib="NCI-CGAP_Sub1"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7AD-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NCI-CGAP_Sub1 library is a subtracted library derived from
 BI. BI constitutes a mixture of 21 normalized or
 subtracted NCI-CGAP libraries: NCI-CGAP_C04,
 NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_C010, NCI-CGAP_C016,
 NCI-CGAP_K1d5, NCI-CGAP_K1d12, NCI-CGAP_K1d3,
 NCI-CGAP_K1d11, NCI-CGAP_Lym2, NCI-CGAP_Br2, NCI-CGAP_C08,
 NCI-CGAP_C1d11, NCI-CGAP_Le12, NCI-CGAP_Brn23, NCI-CGAP_Lu5,
 NCI-CGAP_Lu24, NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6,
 NCI-CGAP_Brn25. These 21 libraries were pooled and a
 single-stranded DNA preparation of the resulting mixture
 was used as a tracer in a subtractive hybridization with a
 driver whose composition is detailed below: NCI-CGAP_K1d3
 pool 1 LAM 3334-3337, 3682-3683, 3798-3803 (IMAGE
 Clonoids 1322376-1323911, 1456008-1456775, 1500552-1502855
) NCI-CGAP_K1d5 pool 1 LAM 3338-3342, 3722-3725,
 3776-3778 (IMAGE Clonoids 1323912-1325831,
 1471368-1472903, 1492104-1493255) NCI-CGAP_Lu5 pool 1 LAM
 3575-3582, 3851-3854 (IMAGE Clonoids 141920-1417991,
 1520904-1522439) NCI-CGAP_GC4 pool 1 LAM 3164-3167,
 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,
 1469064-1470983, 1475592-1476743) NCI-CGAP_Pr22 pool 1
 LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids
 985608-986759, 1101192-1101959, 1217928-1220615)
 NCI-CGAP_C010 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
 Clonoids 1057416-1061255, 1144584-1145351) The resulting
 subtracted library contained 530,000 recombinants.
 Subtraction was performed as previously described [Bonaldo
 & Soares (1996): Normalization and Subtraction:
 Two Approaches To Facilitate Gene Discovery. Genome
 Research 6, 791-806.
 TAG_LIB=NCI-CGAP_Br2
 TAG_TISSUE=breast
 TAG_SEQ=AAACC"

BASE COUNT
 ORIGIN

60 a 125 g 33 t

Query Match 49.1%; Score 303.6; DB 9; Length 333;
 Best Local Similarity 96.1%; Pred. No. 6.5e-43;

Matches 321: Conservative 1; Mismatches 11; Indels 1; Gaps 1;

QY 255 cctgctcgtgtgagttcctcctcagctcagctcgtgtgctcactactact 314
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 333 CTTGCGTATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 274
 QY 315 gggcctcgtccgagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 374
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 273 GGGCGTGGCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 215
 QY 375 ggcgcggtgtcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 434
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 214 GCGGCGGTGTCTTCCGCGCGCGCTCATCAGCTGTGCGGCTTCGTTCA 155

FEATURES

source

QY 435 gggcccgagggcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 494
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 154 GGGCGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 95
 QY 495 ggccttcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 554
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 94 GCGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 35
 QY 555 gggcgtcgc 588
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 34 GGGCTGGCGCTCCGCGCAGGCGTGGCGCTTCGCGGCGG 1

RESULT

9

AM014455 220 bp mRNA linear EST 10-SEP-1999
 LOCUS
 DEFINITION
 UT-H-Bio-aal-a-07-0-UT.s1 NCI-CGAP_Sub1 Homo sapiens cDNA clone
 IMAGE:2709396 3', mRNA sequence.

ACCESSION

AM014455

VERSION

AM014455

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 220)

AUTHORS

NCI-CGAP

TITLE

Unpublished (1997)

JOURNAL

Unpublished (1997)

COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Oligo-dt track not found. Not I site shown in beginning of sequence
 is likely internal to the message. CDNA library preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 Information can be found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA-NO.

FEATURES

source

1. .220
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2709396"
 /clone_lib="NCI-CGAP_Sub1"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7AD-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NCI-CGAP_Sub1 library is a subtracted library derived from
 BI. BI constitutes a mixture of 21 normalized or
 subtracted NCI-CGAP libraries: NCI-CGAP_C04,
 NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_C010, NCI-CGAP_C016,
 NCI-CGAP_K1d5, NCI-CGAP_K1d12, NCI-CGAP_K1d3,
 NCI-CGAP_K1d11, NCI-CGAP_Lym2, NCI-CGAP_Br2, NCI-CGAP_C08,
 NCI-CGAP_C1d11, NCI-CGAP_Le12, NCI-CGAP_Brn23, NCI-CGAP_Lu5,
 NCI-CGAP_Lu24, NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6,
 NCI-CGAP_Brn25. These 21 libraries were pooled and a
 single-stranded DNA preparation of the resulting mixture
 was used as a tracer in a subtractive hybridization with a
 driver whose composition is detailed below: NCI-CGAP_K1d3
 pool 1 LAM 3334-3337, 3682-3683, 3798-3803 (IMAGE
 Clonoids 1322376-1323911, 1456008-1456775, 1500552-1502855
) NCI-CGAP_K1d5 pool 1 LAM 3338-3342, 3722-3725,
 3776-3778 (IMAGE Clonoids 1323912-1325831,
 1471368-1472903, 1492104-1493255) NCI-CGAP_Lu5 pool 1 LAM
 3575-3582, 3851-3854 (IMAGE Clonoids 141920-1417991,
 1520904-1522439) NCI-CGAP_GC4 pool 1 LAM 3164-3167,
 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,
 1469064-1470983, 1475592-1476743) NCI-CGAP_Pr22 pool 1
 LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids
 985608-986759, 1101192-1101959, 1217928-1220615)
 NCI-CGAP_C010 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
 Clonoids 1057416-1061255, 1144584-1145351) The resulting

Subtracted library contained 530,000 recombinants.
Subtraction was performed as previously described [Bonaldo
Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches to Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_LIB-NCI_CGAP_Le12
TAG_TISSUE=leiomyosarcoma
TAG_SEQ=ATATCG

BASE COUNT 34 a 81 c 84 g 21 t
ORIGIN

Query Match 32.4%; Score 200; DB 9; Length 220;
Best Local Similarity 94.1%; Pred. No. 3.5e-25;
Matches 207; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 369 acctatgagcggtgtgtcttcgcgcgcgtcatcagcgtgtgtgcgttcgttcagc 428
|||||
220 ACCATAGCGGGGGTGTCTTCGCCGGGTAAATCAGGTGTGCGCGCTTCGTACGC 161
|||||
Qy 423 gttccgggagcgagggcggtgcgcgcgttcgcgcgttcgttcagcgttc 488
|||||
Db 160 GGTGCGGGGCGCCAGGCGCGATGCGCGCTGCGTGAACGGGCGCCCGCGAGCT 101
|||||
Qy 489 gcttcgcgcgttcgttcgcgcgcgttcgcgcgcgttcgttcagcgttc 548
|||||
Db 100 GGTGCGCGATTTGGCTGTGCGCGCGCTTCGTTCCTTCGTTCGTTCGTTC 41
|||||
Qy 549 agctcgagggcggttcgcgcgcgttcgcgcgcgttcgcgcgcgttcgcgcgc 588
|||||
Db 40 AGCTCGGGGCTGACATCCGCCACGGTGCCTGCGGCGCG 1
|||||

RESULT 10
BF307868 1083 bp mRNA linear EST 21-NOV-2000
LOCUS 601890683F1 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4131836 5',
DEFINITION mRNA sequence.
ACCESSION BF307868
VERSION BF307868.1 GI:11255033
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1083)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
Genomic Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI030 row: e column: 21
High quality sequence stop: 466.

FEATURES

source

Location/Qualifiers
1..1083
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4131836"
/clone_lib="NIH_MGC_17"
/tissue_type="Thadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOT87; Site:1: EcoRI;
Site:2: XhoI; CDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(6). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 218 a 363 c 318 g 184 t
ORIGIN

Query Match 29.3%; Score 181.2; DB 10; Length 1083;
Best Local Similarity 78.2%; Pred. No. 7.1e-22;
Matches 347; Conservative 1; Mismatches 75; Indels 21; Gaps 10;

Qy 63 gatgaacagcagtg-ctgggagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 121
|||||
Db 197 GATGAACAGCAGTGCTGGGAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 256
|||||
Qy 122 ctgcgttcacgt 178
|||||
Db 257 CTCGCTTCAATCGTGCGCGCGCGCTTATGACCTTGTGACCTTGTGCGCTGTG 316
|||||
Qy 179 ccaa-cgttgagcgccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 234
|||||
Db 317 TCATATGTGAGCGCTGTGCTAATGTTTATCTGACCGCGCGCTGTGCGCGCGCG 376
|||||
Qy 235 ctgcgttcacgt 288
|||||
Db 377 CTGCTTCTTACCTGT 436
|||||
Qy 289 ctgcagctgt 347
|||||
Db 437 GCCCGCTGTGCGCTGACACTACCTACCTGCGCGCGCGCGCTGTGCTGTGCTG 496
|||||
Qy 348 ggggacacactactactactactactactactactactactactactactactact 404
|||||
Db 497 GGGGCACTACTACTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 556
|||||
Qy 405 gttgtgcgt 461
|||||
Db 557 GGTGTGCGCTGTGCGCGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 616
|||||
Qy 462 cgtgtgcctacgt 485
|||||
Db 617 CGGTGCTACAGCGCGCACGAGTGC 640
|||||

RESULT 11
AM138178/8 189 bp mRNA linear EST 29-OCT-1999
LOCUS AM138178
DEFINITION UI-H-B11-acy-e-11-0-UI-s1 NCI_CGAP_Sub3 Homo sapiens CDNA clone
IMAGE:2716124 3', mRNA sequence.
ACCESSION AM138178
VERSION AM138178.1 GI:6142578
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 189)
AUTHORS NCI-CGAP http://www.nci.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. CDNA library preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/biopl/image/image.html
Seq primer: M13 Forward
POLYA-No.

FEATURES

source

Location/Qualifiers
1..189
/organism="Homo sapiens"
/db_xref="taxon:9606"

RESULT	12	
CNS0091P		
LOCUS	CNS0091P	925 bp DNA linear
DEFINITION	Drosophila melanogaster genome survey sequence TE3 end of BAC # BaR19b16 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL053013	
VERSION	AL053013.1	GI:49344461
KEYWORDS	GSS.	

RESULT	13
CNS006XK/c	
LOCUS	CNS006XK 935 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BRRI4N09 of Rpcl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL066051

VERSION	KEYWORDS
AL066051.1	GI:4945019
SS.	
SOURCE	Fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 935)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

COMMENT

Web : www.genoscope.cns.it)
determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

FEATURES

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1lb="PRCI-98"
/clone="BACRI14N09"
/note="end : 17"

```

Query Match	11.1%	Score 68.4	DB 12	Length 935
Best Local Similarity	33.1%	Pred. No. 0.015		
Matches 102; Conservative	85;	Mismatches 117;	Indels 4;	Gaps 1

[illegible]

RESULT	14
CNS006XK	
LOCUS	CNS006XK 935 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit

ACCESSION	AL0666051	
VERSION	AL0666051.1	GI:4945019
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila	melanogaster

REFERENCE
AUTHORS
TITLE
JOURNAL

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 935)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage :
Drosophila melanogaster

JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	location/Qualifiers
source	1..935
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_1b="pRC1.98"
	/clone="BACR14N09"
	/note="end : 17"
BASE COUNT	257 a 170 c 162 g 96 t 250 others
ORIGIN	

Query Match	10.9%	Score 67.2	DB 12	Length 935
Best Local Similarity	30.4%	Pred. No. 0.024		
Matches 118, Conservative	103	Mismatches 166	Indels 1	Gaps 1

QY	223	ctggcgacagagccctgctctctactcttcaactgtacactgctcgtgtgtgtagtcttcaac	283
Db	509	CGSGSMKRTSSGSGTCTCCMMYSSVSCGSGSTCTCCGATCGSSCTSCMKSCITYeKKGCG	568
QY	283	ctcaacgtcagcagctgtgtctcaacctaactatgacctgagccctgagacgcgtctccacg	342
Db	569	CGCCTSSSSSSCCCBBSYSTRCSIBCTKCSGCGCTGSGCTGCCGSGSGCGCGSGCGC	628
QY	343	cagcgaggggccaactactacgtgtctcaacctatgcgagcgtgtgtcttcgcgcgcataca	402
Db	629	CGSGGCGCGGCGSSGSGSSGSSGSSGSSGSSGSSGSSGSGSGSGCGCGCGGCGSS	688
QY	403	gcgtgtgtcgcg-tgcggtcttcgtacgcggtgtcccgagccaaagggagtgctccgcgtgcc	461
Db	689	GCGGCGGSGGSGGSGGSGGSGGCGGCGGCGGCGGCGGSGGSGGSGGSGGSGGSGGSGG	748
QY	462	cagtggtcactacgbbgccgcgcgcgcgtgcgtcgtcgtccgttcgctgtcgtgcgcgtgcgc	521
Db	749	SSSSSSCCSGCGCGMSMCSGCGGSSGCGGCCGCGCGSSGCGCGSSGCGCGGCGGSSG	808
QY	522	ctctccctcgtcgtgagacacgcgcgtgtggaagctcgggtgtcgtcctccgcacagtggtctc	581
Db	809	CGGCGCGGSSGSGGSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	868
QY	582	gcgcgcgcctctcgtcgtgtgtcgtcgcg	609
Db	869	CGGSGSGGSGSSGSGSSGCGCGGSGG	896

Mon Jul 15 14:32:24 2002

us-09-899-513-1.rst

Page 11

BASE COUNT 46288 a 33896 c 32934 g 39628 t 201 others
ORIGIN

alignment_scores:
Quality: 1119.00 Length: 205
Ratio: 5.459 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-899-513-2 x AL354876/rev ..

Align seg 1/1 to reverse of: AL354876 from: 1 to: 152947

```

1 MetGlyAlaAlaPheValAlaSerLeuArgSerAsnLeuSerSerAlaThr 17
|||||
54085 ATGGGGGCTGCTTGTGCTAGCTCCGACAGTAACTTTCTTCTGCTCAC 54336
17 rSerArgSerGluMetAsnSerSerValGlyAspLeuGlyValGlyGlyC 34
|||||
54035 TTCAAGCTCAGACATGAACAGCACTGTTGGGAGCTGGTGGCGGCT 53986
34 ySserLeuTPAspAPProAlaArpPheIleValProAlaAlaTyr 50
|||||
53985 GCAAGCTCTGGATGACCTGCTGCTTCACTGATGTCGCGCGCTAT 53936
51 AlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMe 67
|||||
53935 GCCTTGACACAGGGGCTGGGGCTCCAGCCAACTGGGGGCTGGGCAAT 53886
67 tPheIlaArgSerGlyValArgLeuGlyGlnAlaLeuLeuLeuTyrLeuP 84
|||||
53885 GTTCATCCGACAGCGGCGGCGCTGGGCGCAGGCTCTCTACCTGCT 53836
84 heaenLeuAlaLeuValAlaSpGluPhePheThrLeuThrLeuGlnLeuTyr 100
|||||
53835 TCACAGCTGCTGTTGATGATGATTCTTACAGCTCAGCTGACAGCTG 53786
101 LeuThrTyrTyrLeuGlyLeuAlaArgArgProProAlaThrArgProG 117
|||||
53785 CTCACCTACACACTGGGCGCTGGCGGAGCGCGCTGCGCACGGCGGG 53736
117 yProProThrThrcysProProMetArgTyrPheSerSerProArgSers 134
|||||
53735 GCCACCTACACACTGATGATGATGATGATGATGATGATGATGATG 53686
134 eAlaGlyAlaAlaAlaAlaSerTyrAlaValProGlyProGlyArgLeu 150
|||||
53685 CAGCGTGTGCGGCTGGGCTTGTGACGGGCTCCGCGGCGCAGGGCGG 53636
151 ProAlaTrpProGlyAlaTyrGlyAlaProArgAlaLeuProAlaProSe 167
|||||
53635 CCCGCTGGGCGCGGCTGCTACGGGCGCGCGCGGCTGGCGCTTC 53586
167 rProGlyTyrPheArgAlaTrpProLeuProAlaTrpSerThrAlaGlyAla 184
|||||
53585 GCCGTGCTGGGCGGCTGGCGCTGCTGCTGCTGAGACACCGCTGGCG 53536
184 laArgGlyTyrProProArgTrpProSerArgProProProSerCysTrp 200
|||||
53535 CTCGGGCTGGGCTGCTGCGCAGGTGGCTTGGCGCGCTTCCGCTGG 53486
201 CysSerArgProThr 205
|||||
53485 TGCTGGGCGGCAACG 53471

```

seq_name: gb_hhg:AL607089

seq_documentation_block:

LOCUS AL607089 194215 bp DNA linear HTG 30-JAN-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-24819, *** SEQUENCING IN
PROGRESS ****, 2 ordered pieces.
ACCESSION AL607089

VERSION AL607089.15 GI:18476794
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (sites)
JOURNAL Direct Submission

COMMENT Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced g1:17402315.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba24819
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 192819 bases at least Q40
Consensus quality: 193115 bases at least Q30
Consensus quality: 193255 bases at least Q20
Insert size: 194215; sum-of-ctrls
Insert size: 193590; 7.2% error; agarose-fp
Quality coverage: 9.54x in Q20 bases; sum-of-ctrls Quality
coverage: 9.62x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 16618: contig of 16618 bp in length
* 16619 16718: gap of 100 bp
* 16719 194215: contig of 17497 bp in length.
Location/Qualifiers

1. 194215
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-24819"
/clone_lib="RP11-11.1"
1. 194215
/note="assembly_fragment:05205
clone_end:sp6
vector_side:left
clone_end:t7
vector_side:right"

BASE COUNT 55302 a 40938 c 40961 g 56913 t 101 others
ORIGIN

alignment_scores:
Quality: 1119.00 Length: 205
Ratio: 5.459 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-899-513-2 x AL607089/rev ..

Align seg 1/1 to reverse of: AL607089 from: 1 to: 194215

```

1 MetGlyAlaAlaPheValAlaSerLeuArgSerAsnLeuSerSerAlaThr 17
|||||

```

REFERENCE	TITLE	LOCUS	SEQUENCE	ORGANISM	DATE
16201	ARGGGGGGCGCTTTGTGCTAGCCTCCGCACTAACCCTTCTTCGGCAC				16152
17	rserrargserglumetashenserServalcylspleuglyValglyc				34
16151	TTTCAGAGTCAGGAATGAACACGACGTTGGGGACCTCGTGTGGGGGCT				16102
34	ysSerLeuTrpaspaspProalaArgPheIleValValProalaArg				50
16101	GCAGCCTCGGGATGACCTGCTCGCTTCATCTGTGGTCCCGCGGCTAT				16052
51	AlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaLeu				67
16051	GGCTTGGCAGCTGGGCTGGGCGCTGCCAGCCACAGTGGCGGCTGGCAAT				16002
67	pheIleArgserGlyArgLeuGlyAlaLeuLeuLeuLeuLeuLeuLeu				84
16001	GTTTCATCCGACAGCGGGCGCTGGCCAGGACCCCTGCTTCTACCTCT				15952
84	heAsnLeuAlaLeuValAspGluPhePheLeuThrLeuGlnLeuTrp				100
15951	TCAACCTGGCTCTGGTTGATGAGTTCTTCAGCCTCAGCGTGGAGCTGTG				15902
101	LeuThrTYTYTYLeuGlyLeuAlaArgArgProProAlaThrArgProG				117
15901	CTCACCTACTACTGCTGGGCTGGCCGCGAGCGCCCTGCACACCGGCGG				15852
117	YProProThrThrCysProProMetArgArgTrpSerSerProArgSers				134
15851	GGCAGCTACTAGTGTCACACTATGCGGGGGTGTCCTTGGCGGCTCAT				15802
134	erAlaCysAlaAlaAlaAlaSerTYrAlaValProGlyProGlyArgLeu				150
15801	CAGCGTGTGCGCCCTGCGGCTGTCTACGCGGCTCCGCGGCGCAAGGGG				15752
151	ProAlaTrpProGlyAlaThrGlyAlaProArgAlaLeuProAlaProbe				167
15751	CGCGGCTGGCCCGGTGCTACAGGGCGCCCGCGGCTGCTGGCGCTTC				15702
167	rProGlyTrpArgAlaTrpProLeuProAlaTrpSerThrAlaGlyGln				184
15701	GGCTGGTGGCGGGCGGCTGGCCCTCCCTGCTCGAGACCGCTGGGCAAG				15652
184	laArgGlyTrpProProProAlaTrpProSerArgProProSerGlyTrp				200
15651	CTCGGGGCTGGGCTCGCTCCGCGAGCGTGCGCTTGGCGGCTTCGCTGG				15602
201	CysSerArgProThr	205			
15601	TGCTCGCGCCCAAG	15587			
seq_name:	qb_pr:AF080214				
seq_documentation_block:					
LOCUS	AF080214	1534 bp	mRNA	linear	PRI 06-AUG-1998
DEFINITION	Homo sapiens	protease-activated	receptor 4	mRNA, complete cds.	
ACCESSION	AF080214				
VERSION	AF080214.1	GI:3396080			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:				
AUTHORS	Mammalia: Eutheria: Primates: Catarrhini: Homiinae: Homo.				
TITLE	1 (bases 1 to 1534)				
JOURNAL	Kahn, M.L., Hammes, S.R., Botka, C. and Coughlin, S.R.				
REFERENCE	Gene and locus structure and chromosomal localization of the				
AUTHORS	protease-activated receptor gene family				
	J. Biol. Chem. (1998) In press				
	2 (bases 1 to 1534)				
	Kahn, M.L., Zheng, Y.W., Huang, W., Bigornia, V., Zeng, D., Moff, S.,				
	Parsee, R.V., Tam, C. and Coughlin, S.R.				
	A dual thrombin receptor system for platelet activation				
	Nature (1998) In press				
	3 (bases 1 to 1534)				

FEATURES	Source	Location/Qualifiers
AUTHORS	Kahn, M.L. and Coughlin, S.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-JUL-1998) CVRI, UCSF, 513 Parnassus, San Francisco, CA 94143-1030, USA	
FEATURES		
Source		1. .1534 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="19" /map="19p12" 3. .1160 /note="G protein-coupled receptor; thrombin receptor; PAR4"
CDS		1. .1534 /product="protease-activated receptor 4" /protein_id="AAC28860.1" /db_xref="GI:3396081" /translation="MMGRLLMPVLVGLFSLSGTQNPSPVYDESGTGGDSTPILP APRGPOVCAANDSDPTELPDSSRALLGPPRIYVGLPANGIALVWL AQAAPRSPMLMNLTADLLALAPRIAYHLRQGRPFQACRLATPALYGM YGSVLLAAVSLDRYALVPLPLARALRGRLALGMAAMWAAALALPLTQRF RLRSDVIVCHDALPLDAQSHMQPMATCLALGCLPLAMLCYATLHTLAASR RTHALTEAVVLASAAVAFVPSNLILLHYLSDPSPAWGNLYGAVPSLALSTLNSC VDFEIIYYVAEFRDKVRAQILFRSPGDIVASKASASGRGMQTHSLIQ"
BASE COUNT	235 a 526 c 479 g 293 t	1 others
ORIGIN		
alignment_scores:		
Quality:	177.50	Length: 235
Ratio:	1.659	Gaps: 13
Percent Similarity:	45.532	Percent Identity: 30.638
alignment block:		
US-09-899-513-2 x AF080214	..	
Align seg 1/1 to: AF080214 from: 1 to: 1534		
40	ProLlaArgPheIleValValProLlaaIaIaIaIaLeuAlaLeuAlaGlyLe 56	
228	CCCACACAG.....CTGCTGCCGCCCTCATGGCGTCTCTGCTGCT 271	
56	UGlyLeuProLlaasnValAlaAlaLeuAlaMetPheIleArgSerGlyG 73	
272	GGGCGTGGCGGCCAAATGGCGCTGTGGCTGTGGCGTGGCACAGGCCAC 321	
73	LYArgLeuGlyGlnAlaLeuLeuLeuTyrLeuPheasnLeuAlaLeuVal 89	
322	CTCGGCTGCTCCACCATGCTG.....CTGATGACCTCGCGACTGCT 365	
90	AspGluPhePheThrLeuThrLeuGlnLeuTyrLeuTyrTyrLeu. 105	
366	GACCTCTGCTGGCCCTGGCGCTGCCGCCGCCGAGTCCGCTACACCTGGC 415	
106GlyLeuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 118	
416	TGGCAGCGCTGGCCCTTCGGGGAGGCGCCGCTGGCGCCGAGGCGCG 465	
119	ProThr.....ThrcysProPrometArgar 127	
466	CACCTATGATGTCACATGATATGCTCAGTCTCTGCTGGCCGCCGTCAC 515	
127	GTP SerSerProArbSerSerAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 143	
516	CTGGATCTGCTACCTGGCCCTGGTGACACCGCTGGCGGCCGCCCTGGC 565	
144	ValProGlyProGlyArgLeu.....ProAlaTyr..... 153	
566	TGGCGGGCGCTGGCCCTTGACCTGTCATGCTGCTGGCTCATAGCGCG 615	
154ProGlyAlaTyrGln 158	
616	CCGCGCTGGACACTGCTGCACATGCAGCGCACACACTTCGGCTGGCG 665	

158 yAlaProArGAlaLeuProAla..... 165
166 CGCTCCGATCGCGTCTGCTGCCATACAGCGCTGCCCTGAGACGACAGC 715
166ProSerProGlyTTPArGAlaTrp..... 173
1716 CTCCCACTGACACCGCGCTTACCTGCTGCGCTGTGGCTGTTC 765
174ProLeuProAlaTr 178
176 TGCCCTGCTGCGCATGCTGCTACGGGGCCACCTGCACACGCTG 815
178 pSerThraGlyAlaArgGly..... 186
1816 GCGGCGACGCGCGCGCTACGGCCAGCGGCTGACCGCAGTGT 865
187 ..TTPProProArGTrpProSerArg...ProProSerCysTrpCys 201
866 GCTGCGCTCGCGCG...TGCGCTTCTGCTGCGCAGCACTGCTGCTGC 912

_name: gb_pat:AR108643

seq_documentation_block: 4895 bp DNA linear PAT 14-FEB-2001
LOCUS AR108643
DEFINITION Sequence 1 from patent US 6111075.
ACCESSION AR108643
VERSION AR108643.1 GI:12824130
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4895)
AUTHORS Xu, W., Presnell, S. R., Yee, D. P. and Foster, D. C.
TITLE Protease-activated receptor PAR4 (ZCHEMR2)
JOURNAL Patent: US 6111075-A 1 29-AUG-2000;
FEATURES
source 1..4895
/organism="unknown"
BASE COUNT 971 a 1573 c 1400 g 951 t
ORIGIN

alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

Alignment_block:

-09-899-513-2 x AR108643 ..

Align seg 1/1 to: AR108643 from: 1 to: 4895

40 ProAlaArGpHeiLeuAlaProAlaTrpAlaLeuAlaLeuGlyLe 56
401 CCCACACAG...CTGCTGCCCGCTCTAGTGGCTGCTGTGTGT 444
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaLeuPheLeuArgSerGlyG 73
445 GGGGCTGCGCGCATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 494
73 1yArgLeuGlyAlaAlaLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
495 CTCGCTGCGCTGCACGATGCTG...CTGATACACTGCGACTGCT 538
90 AspGluPhePheThrLeuThrLeuGlnLeuTyrLeuTyrLeu. Leu. 105
539 GACCTCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 588
106GlyLeuAlaArgArGProProAlaTrpArGProGlyPro 118
589 TGCCAGCGCGCTGCGCTTGGGGAGCGCGCTGCGCGCTGCGCGCGCG 638

119 ProThr.....ThrCysProPrometArgAr 127
639 CACTATGATGATCATGATGATGATGATGATGATGATGATGATGATGAT 688
127 gTrp.SerSerProArGSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
689 CTGATTCCTACCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 738
144 ValProGlyProGlyArgLeu...ProAlaTrp..... 153
739 TGCCGCGCGCTGCGCTTGGACCTGCTGCTGCTGCTGCTGCTGCTGCT 788
154ProGlyAlaTyrG 158
789 CCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 838
158 yAlaProArGAlaLeuProAla..... 165
839 CGCTCCGATCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888
166ProSerProGlyTTPArGAlaTrp..... 173
889 CTCCCACTGACACCGCGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCT 938
174ProLeuProAlaTr 178
939 TGCCCTGCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 988
178 pSerThraGlyAlaArgGly..... 186
989 GCGGCGACGCGCGCGCTACGGCCAGCGGCTGACCGCAGTGT 1038
187 ..TTPProProArGTrpProSerArg...ProProSerCysTrpCys 201
1039 GCTGCGCTCGCGCG...TGCGCTTCTGCTGCGCAGCACTGCTGCTGC 1085

seq_name: gb_pr:AF055917

seq_documentation_block:

LOCUS AF055917 4895 bp mRNA linear PRI 07-JUL-1998
DEFINITION Homo sapiens protease-activated receptor 4 mRNA, complete cds.
ACCESSION AF055917
VERSION AF055917.1 GI:3293321
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 4895)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Xu, W., Andersen, H., Whitmore, T. E., Presnell, S. R., Yee, D. P.,
Ching, A., Gilbert, T., Davie, E. W. and Foster, D. C.
TITLE Cloning and characterization of human protease-activated receptor 4
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (12), 6642-6646 (1998)
MEDLINE 96283984
REFERENCE 2 (bases 1 to 4895)
AUTHORS Xu, W., Andersen, H., Whitmore, T. E., Gilbert, T., Davie, E. W. and
Foster, D. C.

TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Biochemistry, University of Washington, J
wing, NE Pacific Street, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers
1..4895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p12"
176..1333
/note="PAR4: member of family of G-protein coupled
protease activated receptors"
/codon_start=1
/product="protease-activated receptor 4"
/protein_id="AAC25699.1"
/db_xref="GI:3293322"

CDS

repeat_region	variation	/gene="F2RL3" /frequency="0.01" /replace="A" 2656. .2823
	repeat_region	/rpt_family="MTR1C" /rpt_type=dispersed 2757
	variation	/gene="F2RL3" /frequency="0.21" /replace="A" 2824. .3101
	variation	/rpt_family="AluIub" /rpt_type=dispersed 2852
repeat_region	variation	/gene="F2RL3" /frequency="0.01" /replace="A" 3021
	variation	/gene="F2RL3" /frequency="0.03" /replace="A" 3063
	variation	/gene="F2RL3" /frequency="0.02" /replace="A" 3108
	repeat_region	/gene="F2RL3" /frequency="0.01" /replace="C" 3154. .3470
repeat_region	variation	/rpt_family="AluIo" /rpt_type=dispersed 3282
	variation	/gene="F2RL3" /frequency="0.01" /replace="T" 3327
	variation	/gene="F2RL3" /frequency="0.12" /replace="T" 3484. .3788
	repeat_region	/rpt_family="AluSg" /rpt_type=dispersed 3564
variation	variation	/gene="F2RL3" /frequency="0.01" /replace="T" 3761
	variation	/gene="F2RL3" /frequency="0.02" /replace="T" 3794. .4085
	repeat_region	/rpt_family="MTR1C" /rpt_type=dispersed 4202
	variation	/gene="F2RL3" /frequency="0.01" /replace="G" 4587
variation	variation	/gene="F2RL3" /frequency="0.01" /replace="C" 4611
	variation	/gene="F2RL3" /frequency="0.21" /replace="T" 4753
	variation	/gene="F2RL3" /frequency="0.04" /replace="A" 4765
	variation	/gene="F2RL3" /frequency="0.07"

```
repeat_region      /replace="T"  
                    4988..5133  
                    /pt_family="L1MD1"  
variation          /pt_type=dispersed  
                    4995  
                    /gene="FZRL3"  
                    /frequency="0.35"  
                    /replace="T"  
variation          5255  
                    /gene="FZRL3"  
                    /frequency="0.22"  
                    /replace="C"  
variation          5980  
                    /gene="FZRL3"  
                    /frequency="0.03"  
                    /replace="A"  
variation          6236  
                    /gene="FZRL3"
```

alignment_scores:		
Quality:	177.50	235
Ratio:	1.659	13
Percent Similarity:	45.532	Percent Identity: 30.638

```
alignment_block:
US-09-899-513-2 x AF384819 .
```

Align seg 1/1 to: AF384819 from: 1 to: 11828

[illegible]

```

174 .....Proleuprolaetr 178
      ||| ||
2244 TGCCCTGCTGGCCATGCTGCTGCTACGGGGCCACCTGCGACACGCTG 2293
      |||
178 pSerThrAlaGlyClnAlaArgLy.....186
      |||:: ||::
2294 GCGGCGCCAGCGCGGCTACGGCCACGGCGTGAAGCTGACCGCAGTGGT 2343
      |||
187 ...TrrProProProAargTrrProSerArg...ProProSerCysTrrPcys 201
      ||||| ||||| ||| |||||
2344 GCGGCGCTCGCGCG...TGCGCTTCTTCTGTCGCCACACACCTGCTCTGCG 2390

seq_name: gb_htg:AC008737

seq_documentation_block:
  LOCUS AC008737 248680 bp DNA linear HTG 26-JUL-2001
  DEFINITION Homo sapiens chromosome 19 clone CTD-253869, *** SEQUENCING IN
  PROGRESS ***, 2 ordered pieces.
  ACCESSION AC008737
  VERSION AC008737.9 GI:15021998
  WORDS HTG: HTGS_PHASE2; HTGS_ACTIVERIN.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
  REFERENCE 1 (bases 1 to 248680)
    DOE Joint Genome Institute and Stanford Human Genome Center.
  JOURNAL Direct Submission
  REFERENCE 2 (bases 1 to 248680)
    DOE Joint Genome Institute.
  TITLE Direct Submission
  AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
    Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
    On Jul 26, 2001 this sequence version replaced gi:13752640.
  COMMENT * NOTE: This is a 'working draft' sequence. It currently
    * consists of 2 contigs. Gaps between the contigs
    * are represented as runs of N. The order of the pieces
    * is believed to be correct as given, however the sizes
    * of the gaps between them are based on estimates that have
    * been provided by the submittor.
    * This sequence will be replaced
    * by the finished sequence as soon as it is available and -
    * the accession number will be preserved.
    * 1 155093: contig of 155093 bp. in length
    * 155094 155193: gap of unknown length
    * 155194 248680: contig of 93487 bp in length.
    -----Genome Center
      Center: Joint Genome Institute
      Center Code: JGI
      Web site: http://www.jgi.doe.gov
    -----
      Project Information
      Center Project Name: 804726
      Center clone name: CTD-253869
    -----
      Summary Statistics
      Consensus quality: 247465 bases at least Q40
      Consensus quality: 247949 bases at least Q30
      Consensus quality: 247983 bases at least Q20
      Estimated insert size: 248980; agarose-fp estimation
      Estimated insert size: 248035; sum-of-contigs

  estimation
    Quality coverage: 11.93 in Q20 bases; agarose-fp

  estimation
    Quality coverage: 11.97 in Q20 bases; sum-of-contigs

  estimation
    Location/Qualifiers
      1..248680
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="19"
        /clone="CTD-253869"

```

[illegible]

LOCUS	HSU91939	1400 bp	DNA	Linear	PRI 22-MAR-1997
DEFINITION	Human putative G protein-coupled receptor (GPR25) gene, complete cds.				
ACCESSION	U91939				
VERSION	U91939.1				
KEYWORDS	GI:1905877				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1400)				
AUTHORS	Jung,B.P., Nguyen,T., Kolakowski,L.F. Jr., Lynch,K.R., Heng,H.H., George,S.R. and O'Powd,B.F.				
TITLE	Discovery of a novel human G protein-coupled receptor gene (GPR25) located on chromosome 1				
JOURNAL	Biochem. Biophys. Res. Commun. 230 (1), 69-72 (1997)				
MEDLINE	97148573				
REFERENCE	2 (bases 1 to 1400)				
AUTHORS	Jung,B.P., Nguyen,T., Kolakowski,L.F. Jr., Lynch,K.R., Heng,H.H., George,S.R. and O'Powd,B.F.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-FEB-1997) Pharmacology, University of Toronto, 8 Taddle Creek Rd., Toronto, ON M5S 1A8, Canada				
FEATURES	Location/Qualifiers				
source	1..1400				
gene	/organism="Homo sapiens"				
gene	/db_xref="taxon:9606"				
gene	/chromosome="1"				
gene	/map="1q32.1"				
gene	80..1162				
gene	/gene="GPR25"				
gene	80..1162				
gene	/gene="GPR25"				
gene	/codon_start=1				
gene	/product="putative G protein-coupled receptor"				
gene	/protein_id="AA05015.1"				
gene	/db_xref="GI:1905878"				
gene	/translation="MAPTEWSPSPGAPWDYSGLDGLETELCPAGDLPGYVYIPRA LYLAFAVGLGNFVVMVLLAGRGPRRLVDTEVLHAAADLGVLTLPLMAAAR PWPGDGLCKLSTFALGATSGAGLLAGMSVDVRYLAIVKLELRPLRTPCVASC GMAVVDALGLPSLVYRGLOLPDGGDSSCGEESHAFQGLSILLITLPLVPLV PCVCRISRRRLRRPHVGRARARNSRIIFAIESTVGSMLPFSARAYTHLARLALP PCPLILARMLGTITATLCAVNSCANPULITLLDRSFARALDCACGTGLARRISS ASSISRDSSVFERKRAADANPASASW"				
BASE COUNT	154 a 524 c 474 g 248 t				
ORIGIN					
alignment_scores:					
Quality:	171.00				
Ratio:	1.583				
Percent Similarity:	50.943				
Percent Identity:	32.547				
alignment_block:					
US-09-899-513-2 x HSDU91939	..				
Align seg 1/1 to: HSDU91939 from: 1 to: 1400					
24 SerSerValGlyAspLeuGlyValGly..C 34					
118 AGCGCCCTGGACTACTCGGGGTGGACGCGCAGAGAGCTGGAGCTGT 167					
34 ySerLeuTrpAspAspProAlaIarGPhelIleValValProAlaIaIarY 50					
168 GTCCGGCCGGGACCTCGCCCTACGCTACGTCACTACCTCCCGGCGCTAC 217					
51 AAlaLeuAlaLeuGlyLeuPProAlaAsnValAlaAlaLeuAlaLeu 67					
218 CTGGCGGCGCTCGCCCTGGCGCTGCTGGCGCAC.....GCCTTTGTGGT 261					
67 tPheIleIarGserGlyIArgLeuGlyGlnAlaLeuLeu..LeuTyrL 83					
262 GTGGCTGCTGGCGCGCGCGCGCGCGCGCGCTGTGTGATACCTTGTG 311					

```

83 euphensienualaLeuValaspGluPheThrLeuGlnLeu 99
      :::::::::::::::::::: |||:::||||||| |||
312 TcGTGACCTGGCGGACGTGACCTGCGCTGTGTGTCTACGCTGCCGCTG 361
      TcGTGACCTGGCGGACGTGACCTGCGCTGTGTGTCTACGCTGCCGCTG
100 TrPLeuThrTyrTyrLeuGlyLeuAlaArgArgProProAlaThrArpR 116
      |||::: ||| ||| ||| ||| ||| ||| |||
362 TGGGCGCGG.....GCGGCGGTAGCGGCGCGT..... 390
      TGGGCGCGG.....GCGGCGGTAGCGGCGCGT.....

116 OGlyProProThrThrCysProProMetLArgArgTyrPseSerProArgS 133
      .....GCCGTTGGCGGATGG.....CCCTTGC 412
391 .....
133 eSerAlaCysAlaAlaAlaAlaSerTyrAlaValProGly..... 146
      ::::: |||:::|||||::: |||
413 AAGCTCAGACGCTCGCGCTGGCGGACGCCCTCGCGGCGCGCTGCT 462
147 .....ProGlyArgLeuProAlaThrPr 154
      ||| ||| ||| ||| ||| ||| |||
463 GCTGGCGGCGCATGAGCGTGACCGCTACCTGCGCCGTGGTGAAGCTGCTG 512
154 OGly..AlaTyrGlyAlaProArgAlaLeuProAlaPro..... 166
      |||::: ||| ||| ||| ||| ||| |||
513 AAGCGGAGCGCTGCGGACCCGCGCTGCGCGCGTGGCGCTGCTGCGGC 562
167 .SerProGlyTyrArgAlaTrpPro...LeuProAlaTyrPseThrAlag 182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
563 GCTGTGGCGCGTGGCGCTGCTGCGCGCGCTGCCCTCCCTGTATACGGGG 612
182 Ty'.....GlnAlaArgGlyTrpProPro 190
      |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
613 GTTGCAGCCCTGCTGCGGGCGCAGACAGCAGTGCAGTGGCGAGACCT 662
191 ArgTyrProSerArgProProSerCysTyrCys 201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
663 CCCACGCCCTTCCAGGCGCTCAGCTTGCCTGCTGC 695

seq_name: gb_pr:HUMCHRA

seq_documentation_block:
LOCUS      HUMCHRA               1863 bp      mRNA      linear      PRI 01-NOV-1994
DEFINITION Human Chromogranin A mRNA, complete cds.
ACCESSION  J03483
VERSION    J03483.1 GI:180526
KEYWORDS   Chromogranin A.
           Human pheochromocytoma, cDNA to mRNA, (Library of A. Lamouroux and
           J.Mallet), clone hCGA/42.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 1863)
AUTHORS   Konecki,D.S., Benedum,U.M., Gardes,H.H. and Huttner,W.B.
TITLE     The primary structure of human chromogranin A and pancreasstatin
JOURNAL   J. Biol. Chem. 262 (35), 17026-17030 (1987)
COMMENT   88059106
           Printed copy of sequence for [1] kindly provided by Konecki,D.S.,
           12/15/87.

FEATURES             Location/Qualifiers
     source            1..1863
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
     mRNA              1..1863
                       /map="14q32"
                       /gene="CHGA"
                       /product="chromogranin A mRNA"
     gene              1..1863
                       /gene="CHGA"
     CDS               83..1456
                       /gene="CHGA"
                       /note="Chromogranin A precursor"
                       /codon_start=1
                       /protein_id="AAA52017.1"
                       /db_xref="GI:180527"

```

```

/db_xref="GDB:600-119-777"
/translation="MRSAAVLLALLCAGQVTAIPVNSPMNKSDTEVMKCIYEIVSDPTL
SKSPMPVSOCEETLRGDETLIRHONLKELODLAOGAKERAHQOKHSGED
ELSEVLNENSOAELEKEAVEERSKDYMERKDSKAEKSGEATDGAARQALPEPMO
SKAEENNOAPRGEEREEERATNHPASIPSKYGRPAEGSEGLSOGLYRREKLS
APRGWAKRHEEREEEREEEAEEAEVPEEGFTVLANPHSLGKIKRGESEELA
VGDAGRPGAEADPEEGEESQOKEEEDNMAVPOGLFSGKSGLEGEERLSK
EMEDSRMSKMDOLAKELTAERKRLGEGEEDNRDSMKLSFRARAYGRGPGPOLRR
GWRPSREDLSLGLPLIAYRGYPERKEEGESANRPEDQELSLAISALEKVAHQ
LOCALRG"
sig_peptide
83..136
/gene="CHGA"
/feature="Chromogram in A signal peptide"
137..1453
/gene="CHGA"
/mat_peptide
/product="chromogram in A"
BASE COUNT 437 a 535 c 608 g 283 t
ORIGIN Unreported.

```

```

alignment_scores:
Quality: 168.50 Length: 219
Ratio: 1.793 Gaps: 12
Percent Similarity: 42.922 Percent Identity: 31.963

```

alignment_block:

US-09-899-513-2 x HUMCHRA/rev ..

Align seg 1/1 to reverse of: HUMCHRA from: 1 to: 1863

```

19 ArgSerGluMetAsnSerSerValGlyAspLeuGly.....SerLeuTPA 38
1627 AGGAGTAGAGAACAAAGCGCGTAGGGGCGATAGGGTGGCTGGGCTTGG 1578
31 ValGlyGlyCys.....SerLeuTPA 38
1577 GCAGGCTGGAGGCTGCTCCCTACCGGAGAGCGCTCGGCGCATCTGCGC 1528
38 SPASPRALARGPheIleValProAlaAlaTyrAlaLeuAlaLeu 54
1527 CAGGACCTGCCAAGGAGGACAGCAGGCGGCGCAGGT...GCCCTG 1481
55 GlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSe 71
1480 GGGCGAGCGCTGCCAGCGGTGCTCACGCC...GCCGTAGTGGCTGCAGC 1432
71 rGlyGlyArgLeuGlyGlnAlaLeuLeuLeuTyrLeuPheAsnLeuAlaL 88
1431 TGGTGGCGCCACTTCTCCAGCTGCTT..... 1404
88 euValAspGluPhePheThrLeuThrLeuAlaLeuTyrLeuThr..... 102
1403 .....CAATGGCGGAGCAGGCTCT 1386
103 .....TyrTyrLeuGlyLeuAlaArgArgProAlaAlaTh 114
1385 CCAGCTCCTGCTCTGTGCTGCGTGTGCGCTGCCCTCTCTCTTTC 1336
114 rArgPro.....GlyProProThrThrnc 122
1335 TTCCTCTGGGGTAGCCTCGAGCCTGAGGCGGCGGCGCTCAAGGCT 1286
122 ySPProMetLeuArgTrpSerSerProArgSerSerAlaCysAlaAla 138
1247 GCGGCGCAGGCGCTGAGAGCGGTAGCGCGGCGGAGAGAGAGAGCTTC 1198
1285 GTCTCCCGGGAGATG.....GCCTCCAGCTCTGCTCCAGCT 1248
139 AlaAlaSer.....TyrAlaValProGlyArgLeuProAla 152
1247 GCGGCGCAGGCGCTGAGAGCGGTAGCGCGGCGGAGAGAGAGAGCTTC 1198
132 aTrp.....ProGlyAlaTyrGlyAlaProArgAlaLeuProAla.... 165
1197 ATGGAAGTGTCCCGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1148

```

```

166 .....ProSerProGlyTyrParAlaTrpPro.....LeuProAlaTrp 178
1147 CTCAGCGCTAGCTCTCTGCGCAGCTGCTCATCTGCTCCAGCGTTCG 1098
179 SerThrAlaGlyGlnAlaArgGlyTyrProProArgTrpProSerAr 195
1097 AGT.....CCTCCAGCTCTCTGAGAGCGCGC 1072
195 gProPro 197
1071 TCCTCTCT 1065
seq_name: gb_hhg:AC027113

```

```

seq_documentation_block:
LOCUS AC027113 159391 bp DNA linear HTG 25-MAY-2000
DEFINITION Homo sapiens chromosome 5 clone RP11-67212 map 5, WORKING DRAFT
SEQUENCE, 20 unordered pieces.
ACCESSION AC027113 GI:8081286
VERSION AC027113.3
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 159391)
Homo sapiens chromosome 5, clone RP11-67212
Unpublished
2 (bases 1 to 159391)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barne,N., Bahtien,V., Beda,F.,
Boguski,K., Bouckhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choape,I., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McKean,P., McGurk,A., McKernan,K., McPeckers,R.,
Melchior,J., Menus,L., Mihova,T., Miranda,C., Mienege,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced g1:7656748.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8843
Center clone name: 672-1-2
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149154 bases at least Q40
Consensus quality: 154165 bases at least Q30

```

Consensus quality: 156049 bases at least Q20
Insert size: 195000; agarose-fp
Insert size: 157491; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs

* NOTE: This is a working draft sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1373:	contlg of 1373 bp	in	length
*	1374	1473:	gap of 100 bp	in	length
*	1474	3371:	contlg of 1898 bp	in	length
*	3372	3471:	gap of 100 bp	in	length
*	3472	5991:	contlg of 2520 bp	in	length
*	5992	6091:	gap of 100 bp	in	length
*	6092	8001:	contlg of 1910 bp	in	length
*	8002	8101:	gap of 100 bp	in	length
*	8102	10432:	contlg of 2351 bp	in	length
*	10453	10532:	gap of 100 bp	in	length
*	10553	13499:	contlg of 2947 bp	in	length
*	13500	13599:	gap of 100 bp	in	length
*	13600	17300:	contlg of 3701 bp	in	length
*	17301	17400:	gap of 100 bp	in	length
*	17401	23167:	contlg of 5767 bp	in	length
*	23168	23267:	gap of 100 bp	in	length
*	23368	27394:	contlg of 4127 bp	in	length
*	27395	27494:	gap of 100 bp	in	length
*	27495	31941:	contlg of 4447 bp	in	length
*	31942	32041:	gap of 100 bp	in	length
*	32042	36867:	contlg of 4826 bp	in	length
*	36868	36967:	gap of 100 bp	in	length
*	36968	43237:	contlg of 6270 bp	in	length
*	43238	43337:	gap of 100 bp	in	length
*	43338	52795:	contlg of 9458 bp	in	length
*	52796	52895:	gap of 100 bp	in	length
*	52896	61446:	contlg of 8551 bp	in	length
*	61447	61546:	gap of 100 bp	in	length
*	61547	68822:	contlg of 7276 bp	in	length
*	68823	68922:	gap of 100 bp	in	length
*	68923	79907:	contlg of 10985 bp	in	length
*	79908	80007:	gap of 100 bp	in	length
*	80008	91003:	contlg of 10996 bp	in	length
*	91009	91103:	gap of 100 bp	in	length
*	91104	109448:	contlg of 18325 bp	in	length
*	109429	109528:	gap of 100 bp	in	length
*	109529	135277:	contlg of 25749 bp	in	length
*	135278	135377:	gap of 100 bp	in	length
*	135378	159391:	contlg of 24014 bp	in	length

FEATURES
source

```
misc_feature |>
  mutate(
    /organism="Homo sapiens"  

    /db_xref="taxon:9606"  

    /chromosome="5"  

    /map="5"  

    /clone="RP11-672I2"  

    /clone_id="RPC1-11 Human Male BAC"  

    .1.373  

    /note="assembly-fragment"  

    1474. .3371  

    /note="assembly-fragment"  

    3472. .5991  

    /note="assembly-fragment"  

    6092. .8001  

    /note="assembly-fragment"  

    8102. 10452  

    /note="assembly-fragment"  

    10535. .13495  

    /note="assembly-fragment"
```

BASE COUNT	43727 a	35737 c	36282 g	41744 t	1901 others
misc_feature	13600.	.17300			
	/note="assembly_fragment"				
misc_feature	17401.	.23167			
	/note="assembly_fragment"				
misc_feature	23268.	.27394			
	/note="assembly_fragment"				
	clone_end:17				
	vector_side:right				
misc_feature	27455.	.31941			
	/note="assembly_fragment"				
misc_feature	32042.	.36867			
	/note="assembly_fragment"				
misc_feature	36968.	.43237			
	/note="assembly_fragment"				
misc_feature	43338.	.52795			
	/note="assembly_fragment"				
misc_feature	52896.	.61446			
	/note="assembly_fragment"				
misc_feature	61547.	.68822			
	/note="assembly_fragment"				
misc_feature	68923.	.79907			
	/note="assembly_fragment"				
misc_feature	80008.	.91003			
	/note="assembly_fragment"				
misc_feature	91104.	.109428			
	/note="assembly_fragment"				
misc_feature	109529.	.135277			
	/note="assembly_fragment"				
misc_feature	135378.	.159331			
	/note="assembly_fragment"				
	clone_end:sp6				
	vector_side:right				

[illegible]

```
alignment_block:
```

US-09-899-513-2 x AC027113

Align seg 1/1 to: AC027113 from: 1 to: 159391

24 SerservalGlyaspLeuGlyValGly.....C 34

132146 AGCGCCCTGGGACTACTCGGGGTTGGACGGCCTGGAGGAGCTGGAGCTGT 132195

34 ySSerLeuTrpAspProAlaArgPheIleValProAlaAlaTyr 50

132196 GTCCGGCCGGGACCTGCCCTACGGCTACGTTACATCCCGCGCTCTAC 132245

51 AlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMe 67

132246 CTGGCGCCTTCGCCGTGGCCTGCTGGGCAC.....GCCTTGTGCT 13

67 tPheileargSerGlyIyargLeuGlyGlnAlaLeu...LeuTyrl 83
.....

132290 GATGCTGCTGCGCCGCGCGCGCGCGCGCGCTGGTGGATACCTTCG 13

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

116 0C1vDroDProCthwmbwCueDroDProM0tAr0dvr0E0n60r60rDroAr0c 1133

132433 ATGG...CCTCTGCAAGCTCAGCAGCTTCGGCGCTGGCGGCGACGGCGTGC 132479

(secretory protein I) cDNA
J. Biol. Chem. 263 (23), 11559-11563 (1988)

JOURNAL
MEDLINE
88298816

COMMENT
Draft entry and computer-readable sequence for [1] kindly provided
by L.J. Hellman, 31-MAY-1988.

FEATURES
Location/Qualifiers

source

1..1811

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="14q32"

1..1811

/gene="CHGA"

43..1416

/gene="CHGA"

/note="chromogranin A"

/codon_start=1

/protein_id="AA52018.1"

/db_xref="GI:180529"

/translation="MRSAAVLLALLCAGVATLPVNSPMNKGDTEVMKIVEYIDTL

SKSPMPVSKCEFTLRGDERILITLRHONLKLKLODLAOGKREKRAHOKKSGPED

ELSEVLNDSQALFKQAVEPSSKDYMKRDSRPAEKSGATGARGAOLPEPMOF

SKAEGRKQAPGEEEEEERATITTPASIPSKYRPOAKSGESGLSGLVREKGIS

VEPGWAKKEEEEEEEREAAGEAVPEEGPTVNPSPSLGKIRKESKSEALA

VDGAGKPAEEADPEEGKGEHSDQKGEDEEMAVVPOGLFRGKSGLEDEBERLSK

EMEDNRMKMDOLAKELTAERKRLBQEEEDNRDRSMRLSFRANGYGRGPGQLRK

GSRPNSMEDSLBAGLPLQVRYGYPQEKKEEGSANRRPEDELSLSAIELEKVAHQ

LOCALR"

BASE COUNT

ORIGIN

Chromosome 14.

432 a 509 c 599 g 271 t

alignment_scores:

Quality: 164.50

Ratio: 1.750

Percent Similarity: 42.922

Length: 219

Gaps: 12

Percent Identity: 30.594

alignment_block:

US-09-899-513-2 x HUMCHRAA/rev

Align seg 1/1 to reverse of: HUMCHRAA from: 1 to: 1811

19 ArgSerGluMetAsnSerSerValGlyAspLeuGly.....SerLeuTPA 38

1587 AGCAGTAGAGACAAAGCGCGCTAGGCGCGATAGGCTGCGCTGGCTGG 1538

31ValGlyGlyGly.....SerLeuTPA 38

1537 GCAGGCTGGAGGCTGCTCCTACCGGAGGAGCGGCTCCGGGCGCATCTGGC 1488

38 spAspProAlaArgPheIleValAlaProAlaAlaTyrAlaLeuAlaLeu 54

1487 CAGGACCTGCCAAGGAGGACAGACAGAGCGGCGCAAGGCT...GCCCTG 1441

55 GlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgse 71

1440 GGCGCAGCGCGGCGGCGGCTGTCTCAGCCCGCCTGAGCTGCGAGC 1392

71 rGlyGlyArgLeuGlyAlaLeuLeuLeuTyrLeuPheAsnLeuAlaL 88

1391 TGGTGGGCGCACCTTCTCCAGCT..... 1370

88 euValAspGluPhePheThrLeuThrLeuGlnLeuTyrPheThr..... 102

1369CTGCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1346

103TyrTyrLeuGlyLeuAlaArgArgProProAlaTTh 114

1345 CCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 1296

114 rArgPro.....GlyProProThrThrTTC 122

1295 TTCTCTGGGGGTAGCTCGGAGCTGAGGGGCAAGGCGCGCTCAAGGCT 1246

122 ysProProMetArgArgTyrPheSerSerProArgSerSerAlaCysAlaAla 138

1245 GTCTCT.....CCGAGAGGTGGCGCTCGAGGCTC 1217

139 AlaAlaSerTyrAla.....ValProGlyProGlyArg 149

1216 GCCCGAGCTGCGGCCAGAGCCCGTGAAGCGGTAGCCCGGCGCGGGAAG 1167

149 GluProAlaTyrPhe.....ProGlyAlaTyrGlyAlaProArgAlaLeuP 164

1166 GAGAGCCTCAGGAGAACGGTCCCGGTGCTCTCTCTCTCTCTCTCTCTCTCT 1117

164 roAlaProSerProGlyTyrPheArgAlaTyrProLeu.....ProAlaTyrSer 179

1116 CAGCGCTCTTCAGCCCGTACCTCTGCGAGCGGTGCTCACTCTGCTCC 1067

180 ThrAlaGlyAlaAlaArgGlyTyrPhe.....ProProProArgTyrProSerArg 195

1066 AGC.....GATTGGTGTCTCTCCACTCTCTGAGAGCCGC 1032

195 gProPro 197

1031 TCCTCT 1025

seq_name: gb_pr:BC012755

seq_documentation_block:

LOCUS BC012755 1432 bp mRNA linear PRI 20-AUG-2001

DEFINITION Homo sapiens, clone MGC:16126 IMAGE:3627785, mRNA, complete cds.

ACCESSION BC012755

VERSION BC012755.1 GI:15215328

KEYWORDS MGC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (15-AUG-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: madan@systemsbiology.org

Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia

Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAL Plate: 24 Row: m Column: 23

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

FEATURES

source

1..1432

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:16126 IMAGE:3627785"

/tissue_type="Brain, neuroblastoma"

/clone_lib="NIH_MGC_19"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

23..1009

/codon_start=1

US-09-899-513-2 x AAF21431 ..
Align seg 1/1 to: AAF21431 from: 1 to: 1534

```

40 ProAlaArgPheIleValProAlaAlaTyrAlaLeuAlaLeuGlyLe 56
228 CCCACCAGG.....CTGGTCCCGCCCTCTATGAGCTGCTCTGTTGCT 271
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaLeuAlaLeuArgSerGlyG 73
272 GGGGTGCTGCTGGCCATGAGCGCTGGCTGGTGGTGGTGGTGGTGGTGG 321
73 LyArgLeuGlyGlnAlaLeuLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
322 CTCGGCTGCCCTCCACCATGCTG.....CTCATGACCTCCGCACTGCT 365
90 AspGluPhePheThrLeuThrLeuThrLeuThrLeuThrLeuThrLeu 105
366 GACCTCTGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGG 415
106 .....GlyLeuAlaArgArgProProAlaThrArgProGlyPro 118
416 TGGCCAGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTG 465
119 ProThr.....ThrCysProProMetArgArg 127
466 CACTATAGTGCATGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
127 gTTPSerSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
516 CTGGATCCCTCACTGCGCCCTGCTGCAACCGCGCGCGCGCGCGCTGCG 565
144 ValProGlyProGlyArgLeu..ProAlaTTP..... 153
566 TGGCGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGG 615
154 .....ProGlyAlaTyrG1 158
616 CCGCCCTGGCACTGGCCCTGCACACTGCACGCGGCAACCTTCGGCGTGG 665
158 ValAlaProArgAlaLeuProAla..... 165
666 CGCTCCGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 715
166 .....ProSerProGlyTTPArgAlaTTP..... 173
716 CTCACCACTGGCAACCGCGCTTCACCTGCTGGCGCTGGCGCTGTTGCC 765
174 .....ProLeuProAlaTTP 178
766 TGGCCCTGCTGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 815
178 pSerThrAlaGlyGlnAlaArgGly..... 186
816 GCGGCGACCGCGCGCGCTGACGCGCTGAGAGCTGACCGCAGTGCTG 865
187 ..TTPProProArgTTPProSerArg...ProProSerTyrTTPCys 201
866 GCTGGCTCTCCGCG...TGGCTCTCTGCTGGCGCAACCTGCTGCTGCTG 912
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA35309
seq_documentation_block:
ID AAA35309 standard; DNA; 1534 BP.
XX
XX AAA35309;
XX
XX 28-JUL-2000 (first entry)
XX
XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:183.
XX
XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
KW

```

```

KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukemia; lymphoma; carcinoma; metastasis; ss.
XX
XX Homo sapiens.
XX
XX WO200009525-A2.
XX
XX 24-FEB-2000.
XX
XX 03-AUG-1999; 99WO-US17712.
XX
XX 03-AUG-1998; 98US-0095212.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX
XX NYCE JW;
XX
XX WPI: 2000-205971/18.
XX
XX New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX cancers.
XX
XX Disclosure: Page 1330; 1343pp; English.
XX
XX The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytosolic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impaired respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
XX carcinomas, and cancers which may metastasize to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing the
XX bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
XX differ from the previously named sequences. SEQ ID NO:11 to 1680
XX (AAA32323 to AAA33922) are specifically claimed ONs from the present
XX invention. N.B. Sequences given in the disclosure of the present
XX invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.
XX
XX Sequence 1534 BP; 235 A; 526 C; 479 G; 293 T; 1 other:
XX

```

alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

alignment_block:
US-09-899-513-2 x AAA35309 ..

Align seg 1/1 to: AAA35309 from: 1 to: 1534

```

40 ProAlaArgPheIleValProAlaAlaTyrAlaLeuAlaLeuGlyLe 56
228 CCCACCAGG.....CTGGTCCCGCCCTCTATGAGCTGCTCTGTTGCT 271

```

[illegible]

seq_documentation_block:	
ID	AA32747 standard; cDNA; 4895 BP.
XX	
AC	AA32747;
XX	
DT	31-JAN-2000 (first entry)
XX	
DE	Human protease-activated receptor PAR4 cDNA.
XX	
KW	Protease-activated receptor; PAR4; ZCHEMR2; G protein coupled;
KW	cellular signalling; protease; tethered ligand; N-terminal;
KW	proteolysis; thrombin; trypsin; cleavage; hexapeptide; agonist;
KW	antagonist; cellular response; physiological response; clotting pathway
KW	platelet; proliferation; differentiation; mediation;
KW	inflammatory process; vascular injury; chemotaxis; mitogenesis;
KW	growth factor; production; ds.
XX	
OS	Homo sapiens.
XX	

FT	Key	Location/Qualifiers
FT	CDS	176..1333
FT		/*tag= a
FT		/product= "Human protease-activated receptor PAR4"
FT	5'UTR	1..175
FT		/*tag= b
FT	sig_peptide	176..226
FT		/*tag= c
FT	mat_peptide	227..1330
FT		/*tag= d
FT		/product= "Mature human PAR4"
FT	3'UTR	1333..4895
FT		/*tag= e
FT		/note= "GC-rich"

AA MO9950415-A2.
 AN 07-OCT-1999.
 AP 31-MAR-1999; 99MO-US07100.
 AR 01-APR-1998; 98US-0053866.
 AS (ZYMO) ZYMOGENETICS INC.
 AT (UNIW) UNIV WASHINGTON.
 AU Xu W, Presnell SR, Yee DP, Foster DC;
 WP1; 1999-633640/54.
 DR P-PSDB; AAY50135.
 DT Novel protease activated receptor 4, useful for screening for
 PT (ant)agonists for promoting the proliferation and/or differentiation of
 PT platelets and in mediating inflammatory events -

Claim 1; Page 76-80; 85pp: English.

This sequence represents a human protease-activated receptor PAR4 (also referred to as ZCHEMR2) cDNA which was identified from EST (expressed sequence tag) databases with homology to the three known protease-activated receptors (PAR1, PAR2 and PAR3). Protease-activated receptors (PARs) are a subfamily of G protein coupled receptors which are capable of mediating cellular signalling in response to proteases (e.g., 'thrombin'). They are characterised by a tethered peptide ligand at the extracellular N-terminus that is generated by proteolysis. PAR4 is activated by thrombin or trypsin cleavage at Arg47/Gly48, which generates a new N-terminus corresponding to the tethered ligand (a hexapeptide). As agonists of PAR4 are useful for upregulating cellular or physiological responses whereas antagonists are used to downregulate these activities. The PAR4 protein is further useful for dissecting the effects of thrombin or other activating proteases in the clotting pathway from the effects of these proteases at the cellular level. As agonists are specifically useful in promoting the proliferation and/or differentiation of platelets, in mediating inflammatory events, responses to vascular injury, chemotaxis or mitogenesis, and in producing growth factors. Antagonists are useful as research reagents for characterising sites of ligand-receptor interaction.

Sequence 4895 BP; 971 A; 1573 C; 1400 G; 951 T; 0 other;

```

alignment_scores:
  Quality: 177.50
  Ratio: 1.659
  Percent Similarity: 45.532
  Length: 235
  Gaps: 13
  Percent Identity: 30.638

```

```
alignment_block:
US-09-899-513-2 x AA232747 ..
Align seg 1/1 to: AA232747 from: 1 to: 4895
```

40 ProIaArgPheIleValProIaAlaIATyRAlaLeuAlaLeucylLe 56
|||:::|| |:::||||| |||:::|||:::|| ::


```
401 CCCACGAG.....CTGTGCCCCCTCTATGGCTGTCTGCTGGT 444
56 uGlyleuproAlaasnValAlaAlaLeuAlaMetPheIleArgSerGlyg 73
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 GGGGTGCGCCGCAATGGCGCTGGCGCTGGTGGTGGCCACGAGGAC 494
73 lYArgleuGlyGlnAlaLeuLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 CTGGCTGCCCTCCACCATGCTG.....CTGATGAACCTCGCGACTGCT 538
90 AspGluPhePheThrLeuThrLeuGlnLeuTyrLeuTyrTyrLeu 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
539 GACCTCTGCTGGCGCTGGCGCTGGCGGATGCGCTACACCACTGCG 588
106 .....GlyleuAlaArgArgProProAlaThrArgProGlyPro 118
589 TGCCACGCGCTGGCCCTTGGGAGGCGCGCTGGCGCTGGCCAGGCGG 638
119 ProThr.....ThrcysProPromeTarGar 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
639 CACTGTATGTCACATGTATGGCTCAGTGTCTGCTGGCGCGCTCAGC 688
127 gTTrp.SerSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
689 CTGATGCTGCTACCTGGCGCTGGTGACCCGCTGGCGCGCGCGCTGCG 738
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
739 TGCGCGGCGCTGGCGCTGGCTGCTGATGCTGCTGGCTGATGCGCGG 788
154 .....ProGlyAlaTyrGln 158
789 CCGCCCTGGCACTGCCCTGACACTGACGCGGACACCTTCGGCGTGGG 838
158 yAlaProArgAlaLeuProAla..... 165
839 CGCTCCGATCGCTGCTGCTGCTGACGCGCTGCCCTGGACGACAGCG 888
166 .....ProSerProGlyTrpArgAlaTrp..... 173
889 CTCCCACTGGCAACGCGCTTACCTGCTGGCTGGCTGCTGCTTCC 938
174 .....ProLeuProAlaTrp 178
939 TGCCCTGCTGGCCATGCTGCTGCTACGCGGCGACACCTGACACGCTG 988
178 pSerThrAlaGlyGlnAlaArgGly..... 186
989 GCGGCCACGCGCGCGCTACGCGGCTGAGGCTGACCGCAGTGGT 1038
187 ..TrrProProArgTrpProSerArg...ProProSerCysTrpCys 201
1039 GCTGGCTCTCCGCG...TGGCTTCTTCTGCTGCCAGCAACCTGCTGCTGC 1085

seq_name: /SIS1/gcdata/geneseq/geneseqn-emb1/NA2000.DAT.AA250775
seq_documentation_block:
ID AA250775 standard; DNA; 4895 BP.
XX
XX AA250775;
XX AC
XX DT 31-MAY-2000 (first entry)
XX
DE Human protease activated receptor-4 DNA.
XX
XX Human; PAR-4; protease activated receptor;
XX KW antisease molecule; PAR antibody; cytosolic; therapeutic;
XX metastatic tumour cell; placental implantation; invasive cell; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 176..1333
```

```
FT /*tag= a
FT /product= "PAR-4"
XX
XX WO200008150-A1.
XX
XX 17-FEB-2000.
XX
XX 05-FEB-1999; 99WO-IL00079.
XX
XX 07-AUG-1998; 98IL-0125698.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX Bar-Shavit R;
XX
XX WPI: 2000-205706/18.
XX
XX P-PSDB; AAY45036.
XX
XX Treating metastatic tumor cells useful for treating disorders involving
XX placenta implantation in a female comprises administration of an
XX antisease molecule complementary to an RNA sequence of a protease
XX activated receptor protein -
XX
XX Example 3; Fig 11a; 46pp; English.
XX
XX CC The patent discloses a method to treat metastatic tumour cells using
XX CC an antisease molecule comprising a polynucleotide complementary to an
XX CC RNA sequence of a protease activated receptor (PAR) protein, or an
XX CC antibody capable of binding to a PAR protein. The antisease molecules
XX CC with implantation of placenta. The present sequence is a human PAR-4
XX CC DNA used for producing antisease molecules for treating invasive
XX CC cells.
XX
XX 50 Sequence 4895 BP; 971 A; 1573 C; 1400 G; 951 T; 0 other:

alignment_scores:
Quality: 177.50 Length: 235
Ratio: 1.659 Caps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

alignment_block:
US-09-899-513-2 x AA250775 ..

Align seg 1/1 to: AA250775 from: 1 to: 4895
40 ProAlaArgPheIleValValProAlaAlaTyrAlaAlaLeuAlaLeuGlyLe 56
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 CCCACGAG.....CTGTGCCCCCTCTATGGCTGTCTGCTGGT 444
56 uGlyleuproAlaasnValAlaAlaLeuAlaMetPheIleArgSerGlyg 73
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 GGGGTGCGCCGCAATGGCGCTGGCGCTGGTGGTGGCCACGAGGAC 494
73 lYArgleuGlyGlnAlaLeuLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 CTGGCTGCCCTCCACCATGCTG.....CTGATGAACCTCGCGACTGCT 538
90 AspGluPhePheThrLeuThrLeuGlnLeuTyrLeuTyrTyrLeu 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
539 GACCTCTGCTGGCGCTGGCGCTGGCGGATGCGCTACACCACTGCG 588
106 .....GlyleuAlaArgArgProProAlaThrArgProGlyPro 118
589 TGCCACGCGCTGGCCCTTGGGAGGCGCGCTGGCGCTGGCCAGGCGG 638
119 ProThr.....ThrcysProPromeTarGar 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
639 CACTGTATGTCACATGTATGGCTCAGTGTCTGCTGGCGCGCTCAGC 688
127 gTTrp.SerSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
689 CTGAGATGCTACCTGCGCCCTGTGTGACACCGCTGCGGCCCGCCCTGCG 738
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
||||:|||||
739 TGGCCGGCGGCTGCGCCCTTGAGACTCTGATGCTGCTGGCTATGGCCG 788
154 .....ProGlyAlaTyrG1 158
||||:|||||
789 CCGGCTGTGGACATGCGCCCTGACACTGACGCGGACAGACTTCGGCTGGCG 838
158 yAlaProArgAlaLeuProAla..... 165
|||||
839 CGCTCCGATGCGCGCTGTGCTGACATGACCGCTGCGCCCTGACGACAGCG 888
166 .....ProSerProGlyTrpArgAlaTrp..... 173
||||:|||||
889 CTCCCACTGACACCGCGCTTACCTGCTGCGCTGTGGCTGTTCC 938
174 .....ProLeuProAlaTr 178
|||||
939 TGGCCCTGTGCGCATGCTGTGTGCTACGAGGCGCACCTGACACAGCTG 988
178 pSerThrAlaGlyAlaAlaArgly..... 186
||||:|||||
989 GCGGCCGACGCGCGCGCTACGCGCACGCGCTGAGGCTGACCGCAGTGTGT 1038
187 ..TrpProProArgTrpProSerArg...ProProSerCysTrpCys 201
||||:|||||
1039 GCTGGCCTCCGCGCG...TGGCCTTCTTCGTGCCCACAGCACTGCTGTGCTG 1085
```

seq_name: /STDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH43632

seq_documentation_block:

ID AAH43632 standard; cDNA; 4925 BP.

XX AAH43632:

XX 21-JAN-2002 (first entry)

XX Human PAR4 coding sequence.

XX Human: protease-activated receptor; PAR4; N-terminal; activated PAR4;

XX platelet aggregation; inhibition; tumour cell; proliferation; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 191..1348 /tag= a /product= "PAR4" /note= "CDS derived from protein AAB47623"

XX WO200158930-A1.

XX 16-AUG-2001.

XX 06-FEB-2001: 2001WO-US03807.

XX 09-FEB-2000: 2000US-0500646.

XX (ZYMO) ZYMOGENETICS INC.

XX Balindur N, West RR;

XX WPI: 2001-656678/75.

XX P-PSDB; AAB47623.

XX peptides comprising an amino acid sequence are capable of stimulating
XX protease-activated receptor 4 which are useful in diagnosis and therapy
XX e.g. inhibiting tumor cell proliferation and stimulating platelet
XX aggregation -

XX Disclosure; Page 51-58; 84pp; English.

XX This sequence encodes human protease-activated receptor, PAR4. Peptides
CC derived from, or based on, the N-terminal of activated PAR4 (see
CC AAB47624-77), are capable of stimulating PAR4. These peptides may be
CC used to activate PAR4 at lower concentrations than wild type PAR4.
CC These peptides may be used to stimulate platelet aggregation, and for
CC inhibiting tumor cell proliferation.

XX Sequence 4925 BP; 977 A; 1582 C; 1411 G; 955 T; 0 other;

alignment_scores:

Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

alignment_block:

US-09-899-513-2 x AAH43632 ..

Align seg 1/1 to: AAH43632 from: 1 to: 4925

```
40 ProAlaTrpPheLeuValProAlaTrpAlaLeuAlaLeuGlyLe 56
||||:|||||
416 CCGACACAG.....CTGTGCCCCCTTATGGCTGTGCTGTGTGT 459
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheLeuArgSerGly 73
||||:|||||
460 GGGCTGCGCGGCAATGGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 509
73 LyArgLeuGlyAlaLeuLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
|||||
510 CTGGCTGCTCCCTCCACCATGCTG.....CTGATGAACCTGCGACATGCT 553
90 AspGluPhePheThrLeuThrLeuGlnLeuTrpLeuTrpTyrTyrLeu 105
|||||
554 GACCTCTGCTGGCGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 603
106 .....GlyLeuAlaArgProProAlaTrpArgProGlyPro 118
|||||
604 TGGCAGCGGTGGCCCTTGGGAGGCGCGCTGCGCTGCGCACGCGCG 653
119 ProThr.....ThrcysProMetArgArg 127
654 CACTTATGTACATGATGATGCTGCTGCTGCGCGCGCTGACG 703
127 gTrp..SerSerProArgSerSerAlaCysAlaAlaAlaSerTyrAla 143
|||||
704 CTGATGCTACCTGCGCGCTGTGACCGCGCTGCGCGCGCGCGCTGCG 753
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
||||:|||||
754 TGGCCGGCGCTGGCGCTTGGACTGTGATGCTGTGGCTGTGGCTGTGGCG 803
154 .....ProGlyAlaTyrG1 158
||||:|||||
804 CCGGCTGTGGACATGCGCCCTGACACTGACGCGGACAGACTTCGGCTGGCG 853
158 yAlaProArgAlaLeuProAla..... 165
|||||
854 CGCTCCGATGCGCGCTGTGCTGACATGACGCGCTGCGCGCGCGCGCTGCG 903
166 .....ProSerProGlyTrpArgAlaTrp..... 173
||||:|||||
904 CTCCCACTGACACCGCGCTTACCTGCTGCGCTGTGGCTGTTCC 953
174 .....ProLeuProAlaTr 178
|||||
954 TGGCCCTGTGCGCATGCTGTGTGCTACGAGGCGCACCTGACACAGCTG 1003
178 pSerThrAlaGlyAlaAlaArgly..... 186
||||:|||||
1004 CCGGCGACGCGCGCGCTACGGCGCGCTGAGCGCTGAGCGCATGCTGT 1053
```

187 ..TrrProProArgrtProSerArg...ProProSerCysTrpCys 201
|||||
1054 GCGGCGCTCGCGCG...TGCGCTTCTTCGCGCCACCAACCTCTCTCTCG 1100

seq_name: /SID1/gcdata/geneseq/geneseq-emb1/NA2000.DAT:AAF21433

seq_documentation_block:
ID AAF21433 standard; DNA: 6546 BP.

AAF21433:

14-MAR-2001 (first entry)

Human low adenosine antisense oligonucleotide related sequence #3000.

Low adenosine antisense oligonucleotide: phosphorothioate; allergy;
human; airway disorder; bronchoconstriction; lung inflammation;
surfactant depletion; respiratory; bronchodilator; antiinflammatory;
immunosuppressive; antiallergic; analgesic; hypotensive; cytostatic;
respiratory obstruction; pulmonary obstruction; impeded respiration;
surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
pulmonary hypertension; emphysema; pulmonary transplantation rejection;
chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
cancer; ss.

Homo sapiens.

WO200062736-A2.

26-OCT-2000.

24-MAR-2000; 2000WO-US08020.

06-APR-1999; 99US-0127958.

(UYEC-) UNIV EAST CAROLINA.

(NICE/) NYCE J W.

NYCE JM;

WPI: 2000-679539/66.

Low adenosine (A) content antisense oligonucleotides which do not
trigger adenine receptors during metabolism, useful e.g. for treating
cancers and respiratory obstructions -

Disclosure: Page 1416-1418; 1592pp; English.

The present invention describes low adenosine (A) content antisense
oligonucleotides and compositions (I) comprising them. In the antisense
oligonucleotides the A is replaced by a 'universal' or alternative base.
(I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
immunosuppressive, antiallergic, hypotensive and cytostatic activities.
The antisense oligonucleotides and (I) can be used to down-regulate the
expression and/or activity of target polypeptides associated with
lung/respiratory disorders and malignancies, such as stimulating and
activating peptide factors and transmitters, transcription factors,
immunoglobulins and antibodies, antibody receptors, cytokines and
chemokines, endogenously produced specific and non-specific enzymes,
binding proteins, adhesion molecules and their receptors, cytokine and
chemokine receptors, adenine receptors, bradykinin receptors, central
nervous system (CNS) and peripheral nervous and non-nervous system
receptors, CNS and peripheral nervous and non-nervous system peptide
transmitters, defensins, growth factors, vasoactive peptides and
receptors, binding proteins and malignancy associated proteins. The
antisense oligonucleotides may be used in this way to treat disorders
including respiratory obstruction (especially pulmonary obstruction
and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
and/or surfactant hypoproduction which are associated with a disease or
condition selected from pulmonary vasoconstriction, inflammation,
allergies, asthma, impeded respiration, respiratory distress syndrome
(RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary

CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF21434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ

Sequence 6546 BP; 1603 A; 1691 C; 1500 G; 1751 T; 1 other;

alignment_scores:

Quality: 177.50 Length: 235
Ratio: 1.659 Gaps: 13
Percent Similarity: 45.532 Percent Identity: 30.638

alignment_block:

US-09-899-513-2 x AAF21433 ..

Align seg 1/1 to: AAF21433 from: 1 to: 6546

40 ProAlaArgPheIleValProAlaAlaTyrAlaLeuAlaLeuGlyLe 56
2058 CCACACAGG.....CTGGTGGCCGCCCTCTATGGCTGTCTGTGTGT 2101
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGly 73
2102 GGGGCTGCCGCGCCATGGCGTGGCTGTGGTGTGGCCACGACGAC 2151
73 LyrGlyLeuGlyGlnAlaLeuLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
2152 CTGCGCTGCCCTCCACCATGCTG.....CTGATGAACCTCGGCACTGT 2195
90 AspGluPhePhePheThrLeuThrLeuGlnLeuTyrPheTyrLeu 105
2196 GACCTCTGCTGGCCCTGGCGCTGCCCGCGGATGCCATCCACCTGCG 2245
106GlyLeuAlaArgArgProProAlaThrArgProGlyPro 118
2246 TGSCACGCGCTGCGCTTGGGAGGCGCGCTGCCCTGCCACGCGCG 2295
119 ProThr.....ThcysProBromelargar 127
2296 CACTCTATGCTGCATGATGCTGCATGCTGCTGCTGCGCGCGTACG 2345
127 gTTP..serSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
2346 CTGATGCTGCTACTGCGCTGCGGTGCACCCGCTCGGGCGCGCGCTGCG 2395
144 ValProGlyProGlyLyrGlyLeu..ProAlaTrp..... 153
2396 TGCGCGCGCGCTGCGCTTGGACTGTGCATGCTGCTGTGGCTCATGGCG 2445
154ProGlyAlaTyrGly 158
2446 CCGCCCTGGCACTGCCCTGCACATGCTGCAGCGGACAGACCTTCGCG 2495
158 ValAProAGAlaLeuProAla..... 165
2496 CGGTCCGATCGCTGCTGCTGCATGACGCGCTCCCTGACGACACAGCG 2545
166ProSerProGlyTyrParGlyTrp..... 173
2546 CTCCACATGGCAACCGCGCTTACATGCTGCGGTGTGGGTGTTC 2595
174ProLeuProAlaTrp 178
2596 TGCCCTGCTGGCCATGCTGTGCTACGGGCGCACCTGCACACGCTG 2645
178 pSerThrAlaGlyGlnAlaArgGly..... 186
2646 GCGCGCAGCGCGCGCTACGCGCACGCGCTGAGGCTGACCGCATGTGT 2695
187 ..TrrProProArgrtProSerArg...ProProSerCysTrpCys 201
|||||

2696 GCTGCCCTCCGCCG...TGCGCTTCTTCGTGCCAGCACACCTGCTGTC 2742

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA12000.DAT:AAA3311

seq_documentation_block:

ID AAA3311 standard; DNA; 6546 BP.

XX AAA3311;

DT 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:185.

XX Human adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
XX allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
XX antiallergic; antihistaminic; cytosstatic; analgesic; impaired airway;
XX lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
XX respiratory distress syndrome; pain; cystic fibrosis; emphysema;
XX pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
XX cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.

XX WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasocstriction, inflammation, allergies, asthma, hypertension or
XX bronchitis, emphysema, respiratory distress syndrome, ischaemia or
XX cancers -

XX Disclosure: Page 1331-1333; 1343pp; English.

XX The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytosstatic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impaired respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
XX carcinomas, and cancers which may metastasise to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing
XX bronchoconstriction and inflammation. AAA32313 to AAA3312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
XX differ from the previously named sequences. SEQ ID NO:11 to 1680
XX (AAA32323 to AAA33922) are specifically claimed ONs from the present
XX invention. N.B. Sequences given in the disclosure of the present
XX invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.

XX Sequence 6546 BP; 1603 A; 1692 C; 1500 G; 1750 T; 1 other;

alignment_scores: Length: 235
 Quality: 177.50
 Ratio: 1.659
Percent Similarity: 45.532 Percent Identity: 30.638

alignment_block:

US-09-899-513-2 x AAA3311 ..

Align seg 1/1 to: AAA3311 from: 1 to: 6546

40 ProAlaArgPheIleValAlaProAlaAlaTyrAlaLeuAlaGlyLe 56
|||:||||| |||||:|||||:|||||:|||||:
2058 CCCACCAAG.....CTGTCGCCGCCCTCTATGGCTGCTGCTGCTG 2101
56 UGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheAlaArgSerGly 73
:|||||:|||||:||||| |||||:|||||:
2102 GGGCTCGGGCCAAATAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2151
73 LyArgLeuGlyGlnAlaLeuLeuLeuTyrLeuPheAsnLeuAlaLeuVal 89
||||| :|||||:||||| :
2152 CTGGCTGCCCTCCACCATGCTG.....CTGATGAACCTCGCGCATGCT 2195
90 AspartPhePheIleThrLeuThrLeuGlnLeuThrPheThrTyrLeu 105
||| :|||||:||||| :
2196 GACCTCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 2245
106GlyLeuAlaArgArgProAlaThrArgProGlyPro 118
||| :|||||:|||||:||||| :
2246 TGCGCAGCGCTGCGCTGCGCTGCGCGGAGCGCGCTGCGCTGCGCAGCG 2295
119 ProThr.....ThrcysProPrometAlaArg 127
:|: ||| |||
2296 CACCTATGTCACATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2345
127 GTrp.SerSerProArgSerSerAlaCysAlaAlaAlaAlaSerTyrAla 143
||| ||| ||| :|||||:|||||:|||||:|||||:
2346 CTGATGCTGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCGGCGCGCTGCG 2395
144 ValProGlyProGlyArgLeu..ProAlaTrp..... 153
|||||:||||| |||||
2396 TGCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 2445
154ProGlyAlaTyrGln 158
|||:|||||:|||||
2446 CGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 2495
158 ValAlaProAlaAlaLeuProAla..... 165
||||| ||| ||| |||
2496 CGCTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2545
166ProSerProGlyTyrPargAlaTrp..... 173
|||||:|||||:||||| |||||:|||||:
2546 CTGCCACATGCGAACCGCGCTTACCTGCTGCGCTGCTGCTGCTGCTGCTG 2595
174ProLeuProAlaTrp 178
||| |||
2596 TGCCCTGCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2645
178 PserThrAlaGlyGlnAlaArgGly..... 186
|||||:|||||:||||| |||||:|||||:
2646 GCGGCGACCGCGCGCGCTGACGCGCACCGCGCTGACGCGCACGCTGCTG 2695
187 ..TrrProProArgTrrProSerArg...ProProSerTyrTrpPys 201
|||||:||||| ||| :|||||
2696 GCTGCGCTGCGCGG...TGCGCTTCTGCTGCGCAGCAACCTGCTGCTGCTG 2742

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAx87944
seq_documentation_block:

ID AAx87944 standard; cDNA; 1174 BP.

XX AAx87944;

XX 22-NOV-1999 (first entry)
 XX
 DE Human G protein coupled receptor GPR25 cDNA.
 XX
 KW GPR25; G protein coupled receptor; human; cancer; infection;
 XX pain; diagnosis; therapy; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 79..1164
 FT /*tag= a
 XX
 PN MO9945030-A1.
 XX
 PD 10-SEP-1999.
 XX
 XX 01-MAR-1999; 99MO-US04424.
 XX
 PR 16-MAR-1998; 98US-0039608.
 PR 04-MAR-1998; 98US-0034985.
 XX
 XX (SMIR) SMITHKLINE BEECHAM CORP.
 XX
 PI Elshourbagy N, Lane PA, Tsui P;
 DR WPI; 1999-540815/45.
 DR P-PSDB; AAY31751.
 XX
 PT New messenger polypeptides and polynucleotides useful for diagnosis,
 PT prevention and treatment of cancer and HIV infection - and host
 PT cells and antibodies
 XX
 PS Claim 2; Page 29; 42pp; English.
 XX
 CC This is the nucleotide sequence of an isolated polynucleotide
 CC encoding a novel human G protein coupled receptor termed GPR25
 CC (see AAY31751). The polynucleotide can be obtained from a cDNA
 CC library derived from mRNA in cells of human spleen using expressed
 CC sequence tag analysis. Treatment of a subject in need of GPR25
 CC comprises administering GPR25, or a polynucleotide encoding a
 CC polypeptide that effects production of GPR25 in vivo, while
 CC treatment of a subject in need of inhibiting activity or expression
 CC of GPR25 comprises administering an antagonist of GPR25, a nucleic
 CC acid that inhibits expression of GPR25, or a polypeptide that
 CC competes with GPR25 for its ligand, substrate or receptor. Also
 CC claimed are an expression system, host cells, a membrane of the
 CC host cells and preparation of GPR25. GPR25 is useful for diagnosing
 CC and treatment of a disease or susceptibility to a disease by
 CC determining the presence or absence of a mutation in GPR25 nucleic
 CC acid, and/or analysing for the presence or amount of GPR25
 CC expression (claimed). It is also useful for identifying compounds
 CC that stimulate or inhibit GPR25, by measuring binding of the
 CC candidate to GPR25, and analysing signal of GPR25 (claimed).
 CC Expression in vivo is useful in gene therapy for treatment of
 CC disease. Diseases that can be treated or diagnosed include fungal,
 CC protozoal and viral (e.g. HIV) infections, pain, cancer, diabetes,
 CC obesity, anorexia, bulimia, asthma, Parkinson's disease, acute
 CC heart failure, hypotension, hypertension, urinary retention,
 CC osteoporosis, angina pectoris, myocardial infarction, stroke,
 CC ulcer, allergy, benign prostatic hypertrophy, migraine, vomiting,
 CC anxiety, schizophrenia, manic depression, delirium, dementia,
 CC Huntington's disease and Gilles de la Tourette syndrome.
 XX
 SQ Sequence 1174 BP; 116 A; 443 C; 408 G; 207 T; 0 other;

alignment_scores:
 Quality: 167.50 Length: 200
 Ratio: 1.565 Gaps: 10
 Percent Similarity: 53.500 Percent Identity: 34.500

alignment_block:
 US-09-899-513-2 x AAX87944 ..
 Align seg 1/1 to: AAX87944 from: 1 to: 1174
 24 SerSerValAlGlySplLeuGlyValGly.Gly.....C 34
 117 AAGCGCCCTGGACTACTCGGGGTTGGAGCGCTGGAGAGCTGAGCTGT 166
 34 ySerLeuTrpAspAspProAlaArgPheIleValProAlaAlaTyr 50
 167 GTCCGGCCGGGAGCTCCCTACGCTGCTACCTACCTCCCGCTCTAC 216
 51 AlAlaLeuAlaLeuGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAla 67
 217 CTGGCGGCTTGGCGCTGGGCTGCTGGGCAAC.....GCCTTGTGTGT 260
 67 tPheIleArgSerGlyGlyArgLeuGlyGlnAlaLeuLeu...LeuTyrL 83
 261 GTGGCTGCTGGCGGGCGGGCGGGCGGGCGGGCGGTGGATACCTTGG 310
 83 euphAsnLeuAlaLeuValAspGluPhePheThrLeuThrLeuGlnLeu 99
 311 TCTGCACCTGGCGGAGCTGACCTGGCTGTGCTGCTGCTGCTGCTGCTG 360
 100 TrpLeuThrTyrTyrLeuGlyLeuAlaArgArgProAlaThrArgPr 116
 361 TGGGCCCGGGC.....GGCGGCTAGGCGGCGGCGGCGGCTTGGCG 403
 116 oGlyProProThrThrcysProProweLarArgTrpSerSerProArgS 133
 404 ATGG...CCTTGCAAGCTCAGACGCTTGGCGTGGCGGGACGGCTCG 450
 133 erSerAlaCysAlaAlaAlaSerTyrAlaValProGlyProGlyArg 149
 451 GCGGGCGGCTGCTGCTGGCGGCGGCGGCGGCGGCGGCGGCTGCGCT 500
 150 LeuProAlaTrpProGly...AlaTyrGlyAlaProAlaAlaLeu..... 163
 501 GTGAAGCTGCTCGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 550
 164ProAlaProSerProGlyTrpArgAlaTrpPro...LeuProAl 177
 551 CCGGCTGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCT 600
 177 atPserThrAlaGly.....GlnAlaArg 186
 601 CTGGCTTACCGGGGGTTGCAGCCCTGCTGGGGGCGGAGGAGGAGGAG 650
 186 lYTrpProProArgTrpProSerArgProProSerCysTrpCys 201
 651 CGCGAGGAGCCCTCCACGCTTCCAGGGGCTCAGCTGTGCTGCTG 697
 seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA10298
 seq_documentation_block:
 ID AAA10298 standard; DNA; 1464 BP.
 AC AAA10298;
 XX
 DT 03-JUL-2000 (first entry)
 XX
 DE DNA encoding human APJ receptor.
 XX
 KW APJ receptor; HIV-1 coreceptor; SIV; cellular infection;
 XX envelope protein; env; CD4 coexpression; drug development; ds.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 199..1341
 FT /*tag= a
 FT /product= "Human APJ receptor"

XX MO200014220-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 07-SEP-1999; 99MO-EP06553.
 XX
 PR 08-SEP-1998; 98US-0149045.
 XX
 PA (SCHD) SCHERING AG.
 PI Doms R, Faulds D, Hesselgesser JE, Horuk R, Mitrovic B, Zhou Y;
 DR WPI: 2000-256972/22.
 XX P-PSDB; AAY87466.
 PT New recombinant eukaryotic cells coexpressing APJ and CD4 polypeptides,
 XX useful for identifying compounds that modulate interaction between an
 XX HIV virus and an APJ receptor e.g. to develop anti-HIV drugs
 XX
 Claim 1; Fig 9; 67pp; English.

This sequence represents DNA encoding the human APJ receptor. The APJ receptor is an orphan seven transmembrane domain receptor that has been found to function as an efficient coreceptor for cellular infection by a number of HIV-1 and SIV strains. The entry of HIV into cells involves binding of the viral envelope protein (env) to CD4, followed by interaction with a coreceptor. Binding to the coreceptor triggers a conformational change in env that mediates fusion between the viral membrane and the host cell membrane. The invention relates to recombinant eukaryotic cells transformed with a nucleotide encoding an APJ protein and/or a nucleotide encoding the CD4 protein, such that the cells coexpress APJ and CD4. It also encompasses peptides and antibodies which specifically bind to an extracellular domain of APJ. These inhibit membrane fusion between a cell coexpressing the APJ and CD4 proteins and a cell expressing an HIV env protein and thereby inhibits HIV infection of the APJ/CD4-expressing cell. The new recombinant cells provide an important tool for investigating and controlling HIV infection. The cells can be used to identify compounds that modulate interaction between HIV and an APJ receptor, which may be useful in the development of anti-HIV drugs. Contacting cells expressing APJ and CD4 with an APJ binding/ blocking agent (e.g., the antibodies or peptides) may be useful to inhibit HIV infection of such cells, e.g., to treat subjects having an HIV-related disorder associated with APJ expression. Antibodies and peptide fragments can be included in medicines and administered to treat patients (especially foetuses) having, or at risk of developing, an HIV infection or related disorder. The antibodies can also be used to detect cells expressing the APJ receptor, and are useful to diagnose susceptibility to HIV infection. For example, higher APJ levels in central nervous system tissues may indicate an increased risk of neuropathogenesis associated with HIV infection. Transgenic animals which express human CD4 and APJ proteins may provide model systems for the study of HIV infection and for anti-HIV drug development.

Sequence 1464 BP; 256 A; 466 C; 416 G; 326 T; 0 other;

alignment_scores:
 Quality: 166.00 Length: 228
 Ratio: 1.523 Gaps: 9
 Percent Similarity: 47.807 Percent Identity: 27.193

alignment_block:
 US-09-899-513-2 x AAA10298 ..

Align seg 1/1 to: AAA10298 from: 1 to: 1464

11 SerAsnLeuSerSerAlaThrSerArgSerGluMetAsnSerValG1 27
 ||| ||| ||| : : : : : : : : : : : : : : : : :
 184 TCTCCATCCCGACGATGAGAGAGTGTGATTTTGACAACTACTATGCG 233
 27 YAspLeuGlyValGlyGlyCysSerLeuTrpAspAspProAlaArgPheI 44
 :
 :

234 GGCAGACACACAGTCTGAGTGTAGTACACAGACTGGAAATCTCGGGCG 283
 44 LeValValProAlaAlaThrAlaLeuAlaLeuGlyLeuGlyLeuProAla 60
 :
 284 CCGTCATCCCTGCGCATCTACATGTGTGCTTCTCTCGGGACACCGGGA 333
 61 AsnValAlaAlaLeuAlaMetPheIleArgSerGlyGlyArgLeuGlyG1 77
 ||| :
 334 AACGGTCGTGCTCTCGACCGCTGTTCGAGACCGCGGAGAAAGACGG 383
 77 nAlaLeuLeuLeuTrpLeuPheAsnLeuAlaLeuValAspGluPhePhe 94
 :
 384 CTCAGCTGATATCTTCATTCATCTAGCGTGGCGGTGCTGACGACTTCG 433
 94 hLeuThrLeuGlnLeuTrpLeuThrTyTyrTyrLeuGlyLeuAlaArgAr 110
 :
 434 TGGTAGCGCTGGCCCTGTGGGCTACCTACAGGTACGGGAGATGATGCG 483
 110 gPro.....ProAlaThrArgProGlyProProThrThrCysProPom 125
 ||| :
 484 CCCTTGGGACCTTCTCTGCAAGCTCAGCACTCATCTCATCTGTCGCA 533
 125 eTArgArgTrpSerSerProArgSerSerAlaCysAlaAlaAlaAlaSer 141
 :
 534 CA.....TGTAAGCGACGCTTCTGCTGCTACCGGCTCAGCTTCG 574
 142 TyrAlaVal...ProGlyProGlyArgLeuPro..... 151
 :
 575 ACCGTACTGTGCGCATGCGTAGCGCAGTGCATGCGCTGAGGCTG 624
 152AlaTrpProGlyA 156
 625 CGGGTACAGCGGGCGGTGGCCAGCGCACTTCTTGGGCTGCGCGCT 674
 156 lATrGlyAlaProArgAlaLeuProAlaProSerProGlyTyrArg... 171
 :
 675 CTTGGCATGCTCTGATGATGTTTACGACACACCGGGACTGTGAGACA 724
 172AlaTrpProLeuProAlaTrpSerThrAlaG1 182
 725 CCATAGAGGTGACATGATGATGATGATGATGATGATGATGATGATG 774
 182 yGlnAlaArgGlyTrpProProProArgTrpPro.....SerArgProP 197
 :
 775 TCAGAGT.....GGGCTGGGAGGTGGGCTTGGGGTCTGTCAC 815
 197 ro.....SerCysTrpCysSerArgProThr 205
 ||| :
 816 CACCGTGGGCTTGTGTGTCCTTACCATCA 847
 seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.AAV19862
 seq_documentation_block:
 ID AAV19862 standard; DNA: 954 BP.
 XX
 AC AAV19862;
 XX
 DT 21-JUL-1998 (first entry)
 XX
 DE Gallus domesticus chondrocyte protein gene.
 XX
 KW chondrocyte; transition; proliferation; hypertrophy; diagnosis;
 KW therapy; long bone; lower proliferative; upper hypertrophic;
 KW embryonic vertebrae growth plates; arthritis; treatment;
 KW non-union bone defects; ss.
 XX
 OS Gallus domesticus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..954
 FT /tag= a
 FT /product= chondrocyte protein
 XX

XX MN09801468-A1.

XX PD 15-JAN-1998.

XX PF 30-JUN-1997; 97WO-US11311.

XX PR 05-JUL-1996; 96US-0021672.

XX

PX (UYRP) UNIV ROCHESTER.

XX

PT Reynolds PR:

XX

DR WPI; 1998-110224/10.

XX P-PSDB; AAM52285.

New isolated chondrocyte protein - which is involved in the transition of chondrocytes from proliferation to hypertrophy, used to develop products for diagnosis and therapy

Claim 15; Page 61; 85pp; English.

The sequence is that encoding a protein which is selectively expressed in chondrocytes in lower proliferative or upper hypertrophic zones of long bone and embryonic vertebrate growth plates. The protein is involved in the transition of chondrocytes from proliferation to hypertrophy. The product can be used for identifying the occurrence of proliferation or hypertrophy or the transition of chondrocytes from proliferation to hypertrophy in a tissue sample. It can also be used for preventing chondrocytes from transitioning from proliferation to hypertrophy and for inhibiting arthritic progression of articular chondrocytes in a patient. It can also be used for inducing chondrocytes to transition from proliferation to hypertrophy and for treating non-union bone defects.

Sequence 954 BP; 157 A; 330 C; 308 G; 159 T; 0 other;

Alignment scores:
Quality: 162.50 Length: 129
Ratio: 2.621 Gaps: 7
Percent Similarity: 48.062 Percent Identity: 34.109

alignment block:
US-09-899-513-2 x AAV19862 ..

Align seg 1/1 to: AAV19862 from: 1 to: 954

110 ArGPProAlaThrArgProGlyProProthrhmrcysProprometar 126
||| ||::: ||| |||||| | |||||
125 CGCTGCCTCCGGGCCGACCGCCCCGCCGCGTCCC GCCG... 169

126 gArGTPrSerSerProAra...SerSeraIaCysalaAlaAlaLasert 142
::: ||||| ||||||| ::::::: |||||
170 CTGCGCCCGCTGCAGCTCCGCCGCTGTCGCCAC 212

142 yralavalproglyprogylargleuproalatrpro..... Gly 155
||| :||| ||||| ||||| |||||
213 CCcGTACCACCGCGCGCAATCCCGCTGGCCAGCTGCACGCGC 262

156 AlaTYrGYlaLProARgALaleuProLaPProserPro..... 168
||| :||| ||||| ||||| |||
263 GCCTCCGCCGCGCGCGCGCTGCCCGCGCGCGCGCGCGTGCC 312

169 GlyTrPARgalat 173

313 GCCGTGCCAGATGTGTACAGAGAGCTCGACGGCGCGTGGCGCGC 362

173 rProLeuProAlarPrSerThrrAlagLYlnaIaARGLYtIPProPro 189
:::|||:|||| |||
363 GGCGCGCGCGCACATGAACACTGGCGCGCGC..... TGCGCGCC 406

190 Pro..... ArgTPProSerAr 195

```

407 GCTTCCCGCGCTCTTCGCCGCCGCCGCGCTCGGCGCTGCCAGAGC 456
|||||
195 gPro.....ProSerCySTPpCysSerArgProThr 205
|||||
457 TCACAAGCACCCGCTGCTCAGACGCGCGGCGCTTCC 493

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19864

seq_documentation_block:
ID      AAV19864 standard; DNA; 1347 BP.
XX
XX      AAV19864;
XX
XX      21-JUL-1998 (first entry)
XX
XX      Gallus domesticus chondrocyte protein gene.
DE
XX      chondrocyte; transition;; proliferation; hypertrophy; diagnosis;
XX      therapy; long bone; lower proliferative; upper hypertrophic;
XX      embryonic vertebrae growth plates; arthritis; treatment;
XX      non-union bone defects; ss.
XX
XX      Gallus domesticus.
OS
FH      Key      Location/Qualifiers
FH      CDS      1..1347
FT      /*tag= a
FT      /product= chondrocyte protein
FT
XX      MO9801468-A1.
PN
PD      15-JAN-1998.
XX
XX      30-JUN-1997; 97WO-US11311.
PF
XX      05-JUL-1996; 96US-0021672.
PR
XX      (UTRP ) UNIV ROCHESTER.
PA
XX      Reynolds PR;
PI
XX      WPI: 1998-110224/10.
DR
XX      P-PSDB: AAW52286.
XX
XX      New isolated chondrocyte protein - which is involved in the
XX      transition of chondrocytes from proliferation to hypertrophy, used
XX      to develop products for diagnosis and therapy
XX
XX      Claim 19; Pages 64-65; 85pp; English.
XX
XX      The sequence is that encoding a protein which is selectively
XX      expressed in chondrocytes in lower proliferative or upper
XX      hypertrophic zones of long bone and embryonic vertebrae growth
XX      plates. The protein is involved in the transition of
XX      chondrocytes from proliferation to hypertrophy. The product
XX      can be used for identifying the occurrence of proliferation
XX      or hypertrophy or the transition of chondrocytes from proliferation
XX      to hypertrophy in a tissue sample. It can also be used for
XX      preventing chondrocytes from transitioning from proliferation
XX      to hypertrophy and for inhibiting arthritic progression of
XX      articular chondrocytes in a patient. It can also be used for
XX      inducing chondrocytes to transition from proliferation to
XX      hypertrophy and for treating non-union bone defects.
XX
SQ      Sequence 1347 BP; 266 A; 432 C; 392 G; 257 T; 0 other;

alignment_scores:
Quality: 162.50      Length: 129
Ratio: 2.621      Gaps: 7
Percent Similarity: 48.062      Percent Identity: 34.109

```



```

XX 21-JUL-1998 (first entry)
XX
XX Gallus domesticus chondrocyte protein gene.
XX
XX chondrocyte; transition; proliferation; hypertrophy; diagnosis;
XX therapy; long bone; lower proliferative; upper hypertrophic;
XX embryonic vertebrae growth plates; arthritis; treatment;
XX non-union bone defects; ss.
XX
XX Gallus domesticus.
XX
XX Key Location/Qualifiers
XX FH 81.1037
XX CD /tag=a
XX FT /product= chondrocyte protein
XX
XX WO9801468-A1.
XX
XX 15-JAN-1998.
XX
XX 30-JUN-1997; 97WO-US11311.
XX
XX 05-JUL-1996; 96US-0021672.
XX
XX (UYRP ) UNIV ROCHESTER.
XX
XX Reynolds PR;
XX
XX WPI; 1998-110224/10.
XX DR P-SDB; AAM52285.
XX
XX New isolated chondrocyte protein - which is involved in the
XX transition of chondrocytes from proliferation to hypertrophy, used
XX to develop products for diagnosis and therapy
XX
XX Claim 14; Pages 59-60; 85pp; English.
XX
XX The sequence is that encoding a protein which is selectively
XX expressed in chondrocytes in lower proliferative or upper
XX hypertrophic zones of long bone and embryonic vertebrae growth
XX plates. The protein is involved in the transition of
XX chondrocytes from proliferation to hypertrophy. The product
XX can be used for identifying the occurrence of proliferation
XX or hypertrophy or the transition of chondrocytes from proliferation
XX to hypertrophy in a tissue sample. It can also be used for
XX preventing chondrocytes from transitioning from proliferation
XX to hypertrophy and for inhibiting arthritic progression of
XX articular chondrocytes in a patient. It can also be used for
XX inducing chondrocytes to transition from proliferation to
XX hypertrophy and for treating non-union bone defects.
XX
XX
XX Sequence 2233 BP; 534 A; 579 C; 601 G; 519 T; 0 other:
XX
XX
XX alignment_scores:
XX
XX Quality: 162.50 Length: 129
XX Ratio: 2.621 Gaps: 7
XX Percent Similarity: 48.062 Percent Identity: 34.109
XX
XX alignment_block:
XX US-09-899-513-2 x AAV19861 ..
XX
XX Align seg 1/1 to: AAV19861 from: 1 to: 2233
XX
XX 110 ArgProAlaIthArgProGlyProProthThrcysProPometaR 126
XX ||| |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 205 GCGTGTCCCTCGGCGCCGACGCCGCCGCCGCCGCGTCCGCCGCCG... 249
XX
XX 126 gArGtrpserProAyd...SerrrAlaGySaLaAlaAlaAlaSerT 142
XX ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 250 .....CTGCGCCCGCGCTCAGCTCCGCGCGTGTGTCGCGACAGCMA 292

```

```

142 yfAlaValProGlyProGlyLeuProAlaTrpPro..... Gly 155
143 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 CGCTACCCCGCGCGGCAATCCGCGCTGCGAGCTGCACGCGC 342
156 AlaTrpGlyAlaProArgAlaLeuProAlaProSerPro..... 168
157 ||| ::||| ||||| ||||| ||||| ||||| |||||
343 GCCTCGCGCGCGCGCGCGCGCTCTCTCCCGCGCGCGCGCTGCGC 392
169 ..... GlyTrpArgAlaTr 173
170 ||||| ||||| ||||| ||||| ||||| ||||| |||||
393 GCCTGCGCGATGTGTACGAGAGAGCCTGCACGGCGCGTGGCGCGC 442
173 rpProLeuProAlaTrpSerTrpAlaGlyAlaArgGlyTrpPro 189
174 ::||| ||||| ||||| ||||| ||||| ||||| |||||
443 GGGCGCGCGCGCATGGAACACTGGCGCGCGCGC..... TGCGCGCGC 486
190 Pro..... ArgTrpProSerAr 195
191 ||||| ||||| ||||| ||||| ||||| ||||| |||||
487 GGTTCGCCGCGCTCTTCGCGCGCGCGCGCGCGCTTGCGCTGGCCAGCAGC 536
195 gPro..... ProSerCysTrpCysSerArgProThr 205
196 ||| ::||| ||||| ||||| ||||| ||||| |||||
537 TCCAAGCACCGCTGCTGCAGAGCGCGCGCGCTTCC 573

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-embL/MA1998.DAT:AAV19860
seq_documentation_block:
ID AAV19860 standard; DNA: 5027 BP.
XX
XX AAV19860;
XX
XX 21-JUL-1998 (first entry)
XX
XX Gallus domesticus chondrocyte protein gene.
XX
XX chondrocyte; transition;; proliferation; hypertrophy; diagnosis;
XX therapy; long bone; lower proliferative; upper hypertrophic;
XX embryonic vertebrae growth plates; arthritis; treatment;
XX non-union bone defects; ss.
XX
XX Gallus domesticus.
XX
XX OS
XX
XX FT FT Location/Qualifiers
XX CDS 81..1037
XX /*tag= a
XX /product= chondrocyte protein
XX
XX WO9801468-A1.
XX
XX 15-JAN-1998.
XX
XX 30-JUN-1997; 97WO-US11311.
XX
XX 05-JUL-1996; 96US-0021672.
XX
XX (UTRP ) UNIV ROCHESTER.
XX
XX PA
XX PA Reynolds PR;
XX PI
XX PI
XX WIPI: 1998-110224/10.
XX DR P-PSDB; AAW52285.
XX
XX New isolated chondrocyte protein - which is involved in the
XX transition of chondrocytes from proliferation to hypertrophy, used
XX to develop products for diagnosis and therapy
XX
XX Claim 13; Pages 55-58; 85pp; English.
XX
XX The sequence is that encoding a protein which is selectively
XX expressed in chondrocytes in lower proliferative or upper
XX hypertrophic zones of long bone and embryonic vertebrae growth
XX plates. The protein is involved in the transition of
XX chondrocytes from proliferation to hypertrophy. The product

```

CC can be used for identifying the occurrence of proliferation
CC or hypertrophy or the transition of chondrocytes from proliferation
CC to hypertrophy in a tissue sample. It can also be used for
CC preventing chondrocytes from transitioning from proliferation
CC to hypertrophy and for inhibiting arthritic progression of
CC articular chondrocytes in a patient. It can also be used for
CC inducing chondrocytes to transition from proliferation to
CC hypertrophy and for treating non-union bone defects.
XX
SQ Sequence 5027 BP: 1353 A; 1136 C; 1200 G; 1338 T; 0 other;

alignment_scores: Quality: 162.50 Length: 129
 Ratio: 2.621 Gaps: 7
Percent Similarity: 48.062 Percent Identity: 34.109

alignment_block:
US-09-899-513-2 x AAV19860 ..

align seg 1/1 to: AAV19860 from: 1 to: 5027

```
110 ArgProProAlaThrArgProGlyProProThrThrcysProPrometar 126
    ||| |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 CGCTGTCCTCGGGCCGACGCCGCCGCCGCCGCCGCTGCGCCG... 249
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 gArGTrpSerSerProArg...SerSerAlaCysAlaAlaAlaLasert 142
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 .....CTGCGGCCCGCTGCACCTCGCGCGCTGCTCGCCGAC 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 yralaValProGlyProGlyArgLeuProAlaTrpPro.....Gly 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 CCGCTACCCGACGGCGGCAATCCGCGCGCTGCGCGAGCTGCACGGCC 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 AlaTyrglyAlaProArgAlaLeuProAlaProSerPro..... 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 GCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 .....GlyTrpArgAlaTr 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
393 GCCTGGCGCATGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 rProLeuProAlaTrpSerThrAlaGlyAlaAlaArgGlyTrpProPro 189
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
443 GGGCGCGCGAGACATGAGAACACTGGCGCGCGCC.....TGGCGGCC 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 Pro.....ArgTrpProSerAr 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
487 GCTTCCCGCGCTTCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGAC 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 gPro.....ProSerCysTrpCysSerArgProThr 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
537 TCCAAGCACGCGCTGCTGACAGAGGCGGCGGCGCTTCC 573
```



```

|||||
382 ACCTGGCGCCCTGGCAATGTCATCCAGCGGCGGCGCTGGCCAG 333
78 AAlaLeuLeuTyrLeuPheAsnLeuValaAspGluPhePhe 94
332 GCCCTGCTTCTACTCTCAACCTGCTGTTGATGATGATCTTCC 283
94 rLeuThrLeuGlnLeuTyrPheTyrTyrLeuGlnLeuAlaArg 111
282 GCTACGCTGACAGCTGTGCTCCTACTACTGCGCTGGCCGCGAG 233
111 rOPrOAlaThrArgProGlyProProThrThCysProPronet 127
232 CGCCTGCCACGCGCGCGGCGCACCTACTACTGTCACCTATGCG 183
128 TrpSerSerProArgSerSerAlaCysAlaAlaAlaSerTyr 144
182 TGGTCTTGGCGGCTCATACAGCTGTGCTGCGCTCTCTACGCG 133
144 rProGlyProGlyArgLeuProAlaTyrProGlyAlaTyrGly 161
132 CCGGCGCCACAGGCGGCGCTGCCGCTGGCCCTACGCGCGCC 83
161 rGAlaLeuProAlaProSerProGlyTyrPArgAlaTyrPro 177
82 GCGCGCTGCTGCGCTGCTGCGCTGCGCGCTGCGCTGCTGCC 33
178 TrpSerThrAlaGlyAlaAlaArgGlyTyr 187
32 TGGACACCGCTGGGCAAGCTGGGGCTGG 3

```

seq_name: gb_est2:BF195670

seq_documentation_block:

LOCUS BF195670 490 bp mRNA linear EST 03-NOV-2000

DEFINITION 7/86906.x1 NCI-CGAP_Ov18 Homo sapiens CDNA clone IMAGE:3571762 3'

sequence.

ACCESSION BF195670

VERSION BF195670.1 GI:11082804

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

CONTACT: Robert Strausberg, Ph.D.

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emerit-Buck, M.D., Ph.D. CDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL, send email to:

info@image.lnln.gov

High quality sequence stop: 452.

Location/Qualifiers

1. 490

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3571762"

/clone_lib="NCI-CGAP_Ov18"

/tissue_type="fibrocyte"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGGCGCGGACATTTTATTTTATTTT 3'];

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Patricia Bonaldo.

BASE COUNT 94 a 158 c 190 g 48 t

alignment_scores: Quality: 856.00 Length: 158 Ratio: 5.452 Gaps: 0 Percent Similarity: 99.367 Percent Identity: 98.734

alignment_block: US-09-899-513-2 x BF195670/rev ..

Align seg 1/1 to reverse of: BF195670 from: 1 to: 490

```

39 AspproAlaArgPheIleValAlaProAlaAlaTyrAlaLeuAla 55
|||||
488 GACCTGCTGCTTCATGCTGTGCTGCGCGGCTATGCTTGCACTGG 439
55 rLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIle 72
438 CCGGCGGCTGCCACGCAAGCTGACGCGCTGCGCAATGTCATCC 389
72 rGlyArgLeuGlyGlnAlaLeuLeuTyrLeuPheAsnLeuAla 88
388 GCGGCGGCTGGGCCAGGCGCTGCTTCTTACTGTTCACTGGCTTG 339
88 ValAspGluPhePheThrLeuThrLeuGlnLeuTyrTyrTyr 105
338 GTTGAATGAGTTCTTCACGCTCAGCTGAGCTGTGCTCCTACTAC 289
105 uGlyLeuAlaArgArgProProAlaThrArgProGlyProPro 122
288 GCGCGCTGGCGCGGAGCGCGCTGCGCACGCGGCGGCGCACCT 239
122 rProPronetArgArgTyrTrpSerProArgSerSerAlaCys 138
238 GTCCACTATGCGGCGGCTGTGCTTCCGCGCTCATACGCTGCTG 189
138 AlaAlaSerTyrAlaValAlaProGlyProGlyArgLeuProAl 155
188 GCGGCTGTGCTACGCGGCTGCCGCGGCGGCGGCGGCTGCGCG 139
155 ValArgGlyAlaProAlaGAlaLeuProAlaProSerProGly 172
138 TCCCTACGCGCGCGCGCGCGCTGCTGCGCTGCTGCGCGG 89
172 rATrProLeuProAlaTyrTrpSerThrAlaGlyAlaAlaArg 188
88 CCGGCGGCTGCCGCTGCGGAGCACCGCAAGCAAGCTGCGGCTG 39
188 ProProArgTyrTrpProSerArgPro 196
38 CGGCCACGCTGCGCTTCCGCGCGG 15

```

seq_name: gb_est1:A1264420

seq_documentation_block:

LOCUS A1264420 460 bp mRNA linear EST 28-JAN-1999

DEFINITION qk54f04.x1 NCI CGAP C08 Homo sapiens CDNA clone IMAGE:1872799 3'

sequence.

ACCESSION A1264420

VERSION A1264420.1 GI:3872623

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.

REFERENCE 1 (bases 1 to 460)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1000 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 449.
Location/Qualifiers
1. 460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1872799"
/clone_lib="NCI-CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo.
BASE COUNT 87 a 151 c 179 g 42 t 1 others
ORIGIN
alignment_scores:
Quality: 816.00 Length: 153
Ratio: 5.368 Gaps: 0
Percent Similarity: 99.346 Percent Identity: 98.693
alignment_block:
US-09-899-513-2 x AI264420/rev ..
Align seg 1/1 to reverse of: AI264420 from: 1 to: 460
38 AspAspProAlaArgPheIleValProAlaAlaTyrAlaLeuAla 54
|||||
460 GATGACCTGCTGCTTCATCGTGGTCCGCGGCTATGCTTGGCACT 411
54 uGlyLeuGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgS 71
|||||
410 GGGCCCTGGGGCTCCAGCCCAAGCTGNCGGCCCTGCAATGTCATCCGCA 361
71 erGlyGlyArgLeuGlyGlnAlaLeuLeuTyrLeuPheAsnIleAla 87
|||||
360 GCGGGGGGGCTGGGGCGAGCCCTCTCTTACCTGTTCAACCTGGCT 311
88 LeuValAspGluPhePheThrLeuThrLeuGlnLeuTyrLeuThrTyr 104
|||||
310 CTGGTTGATGAGTTCCTTCAAGCTCAGCTGCTGCTGCTACCTACTA 261
104 rLeuGlyLeuAlaArgArgProProAlaThrArgProGlyProProThrT 121
|||||
260 CCTGGGCTGGGCGCCGAGGCGCTGACAGCGCGCGGCGCCACACTACTA 211
121 hrcysProPometaArgArgTyrSerSerProArgSerSerAlaCysAla 137
|||||
210 CGGTGCACTATGCGGGGTGGCTTCGCGCGCTCATAGGGTGGCC 161
138 AlaAlaAlaSerTyrAlaValProGlyProGlyArgLeuProAlaTyrPr 154

|||||
160 GCGGCGCTTCGACGCGGCTCCGCGGCGGCGGCTGCGCCCTGCGCC 111
154 oGlyAlaTyrGlyAlaProAlaArgAlaLeuProAlaProSerProGlyTyr 171
|||||
110 CGGTGCTTACGGGCGCCCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 61
171 rGAlaTrpProLeuProAlaTyrSerThrAlaGlyGlnAlaArgGlyTyr 187
60 GGGCCCTGGCCCCCTCTGCTCTGAAGCACCCTGGCAAGCTCGGGCTGG 11
188 ProProPro 190
|||||
10 CTCGCCCA 2
seq_name: gb_est1:AI277080
seq_documentation_block:
LOCUS AI277080 466 bp mRNA linear EST 29-JAN-1999
DEFINITION q140h02.x1 NCI-CGAP_C08 Homo sapiens CDNA clone IMAGE:1874835 3',
mRNA sequence.
ACCESSION AI277080
VERSION AI277080.1 GI:3899348
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 466)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 978 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 431.
Location/Qualifiers
1. 466
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1874835"
/clone_lib="NCI-CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo.
BASE COUNT 89 a 153 c 177 g 45 t 2 others
ORIGIN
alignment_scores:
Quality: 805.00 Length: 153
Ratio: 5.331 Gaps: 0
Percent Similarity: 98.693 Percent Identity: 98.039
alignment_block:
US-09-899-513-2 x AI277080/rev ..

seq_name: gb_est1:AW014139

LOCUS	AW014139	333 bp	linear	EST 10-SEP-1999
DEFINITION	UI-H-BIO-aaj-b-07-0-UI.S1 NCL_CGAP_Sub1 Homo sapiens cDNA clone IMAGE:2709445 3', mRNA sequence.			

VERSION AW014139.1 GI:58628966
KEYWORDS EST.

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 333)

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

Oligo-dT track not found. Not 1 site shown in beginning of sequence. It is likely internal to the message. cDNA library Preparation: M.B. Scores lab clone distribution. NCI-CGAP clone distribution. Information can be found through the I.M.A.G.E. Consortium/INL at: www-bio.lnl.gov/bbrp/image/image.html The following repetitive elements were found in this cDNA sequence: 236-273, >GC_r1ch#low_complexity Seq primer: M13 Forward POLYA-NO.

FEATURES

Location/Qualifiers

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2709445"
/clone_lib="NCI_CGAP_Sub1"

```

BASE COUNT
ORIGIN

60 a -115 c 125 g 33 t

/lab.host="Bnl10B (Life Technologies)"
"/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1 : Not I; Site.2 : Eco RI; The
NCL_CGAP_SubI library is a subtracted library derived from
BI. BI constitutes a mixture of 21 normalized or
subtracted NCL_CGAP libraries: NCL_CGAP_Gc4,
NCL_CGAP_Pr22, NCL_CGAP_Pr28, NCL_CGAP_Co10, NCL_CGAP_Co16,
NCL_CGAP_Kid5, NCL_CGAP_Kymd12, NCL_CGAP_Kid3,
NCL_CGAP_Kid11, NCL_CGAP_Lyml2, NCL_CGAP_Br2,
NCL_CGAP_Cll1, NCL_CGAP_Le12, NCL_CGAP_Brn23, NCL_CGAP_Lu505,
NCL_CGAP_Cll1, NCL_CGAP_Le12, NCL_CGAP_Gc4, NCL_CGAP_GC6
, NCL_CGAP_Lu24, NCL_CGAP_Lu19, NCL_CGAP_Gc4,
, NCL_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with a
driver whose composition is detailed below: NCL_CGAP_Kid3
pool 1 LLM 3334-3337, 3682-3683, 3798-3803 (IMAGE
Clonoids 1322576-1323911, 1456008-1456775, 1500552-1502855)
) NCL_CGAP_Kid5 pool 1 LLM 3338-3342, 3722-3725,
3776-3778 (IMAGE Clonoids 1333912-1325631,
1471368-1472903, 1492104-1493255) NCL_CGAP_Lu5 pool 1 LLM
3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991,
1520904-1522439) NCL_CGAP_Gc4 pool 1 LLM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clonoids 1257099-1258631,
1465064-1470983, 1475593-1476743) NCL_CGAP_Pr22 pool 1
LLM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids
985508-986759, 1101192-1101959, 1217928-1220615)
NCL_CGAP_Co10 pool 1 LLM 2644-2653, 2871-2872 (IMAGE
Clonoids 1057416-1061255, 1144584-1145351). The resulting
Subtraction was performed as previously described [Bonald
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_Lib-NCL_CGAP_Br2
TAG_TISSUE-breast
TAG_SEQ-AAAC"

[illegible]

alignment_block:
US-09-899-513-2 x AW014139/rev

Align seg 1/1 to reverse of: AM014139 from: 1 to: 333

86 LeuAlaIeuValAspIleuPhehLeuThrLeuThrLeuGlnLeuTrpLeuTh 102
332 CTGGCTGATGGTGAATGATTATTCATGCTAACGCTGCAGACTGTGGCTCAC 283
102 rTyrTyLeuGlyLeuAlaArgArgProProAlaThrArpProGlyProp 119
282 ATACTACTAGGGCTGGCCCCGGGGGGCCGCTTCC .ACGGGGCCGGGGGCAC 233
119 rothrThrCysPropPometaArgArgTpsSerSerProArgSerSera 135
233 ATAAATACGGACACCACTTATGGCGGGGTGGTTCGCCGCTCATACAGC 184
136 CysAlaAlaAlaAlaAlaSerTyAlaValProGlyProGlyArGleuProAl 155
183 TGTGGCCGCTGGCTTCGTACACGGGTCGCCGGGCCAGGAGGCGGCTGCCG 134
152 aTrpProGlyAlaTyrglyAlaProArgAlaIeuProAlaProSerProG 168
133 ATGGCCCGGTGACATAGGGGCCCGCGCGCGCTGGCTGGCTTGGCCGTG 84
169 LyrTrpArgAlaArpProIeuProAlaTrpSerThrAlaGlyGlnAlaArg 185
83 GCTGGGGGGGCTGGCCCTTCCTGGCTGGAGACACCGCTGGGCAAGCTCG 34


```

/clone_lib="NCI_CGAP_Sub1"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pTZ19-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI. The
NCI_CGAP_Sub1 library is a subtracted library derived from
BT. BI constitutes a mixture of 21 normalized or
subtracted NCI_CGAP libraries: NCI_CGAP_Co4,
NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16,
NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3,
NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br23, NCI_CGAP_Co8,
NCI_CGAP_Cu11, NCI_CGAP_Le12, NCI_CGAP_Br23, NCI_CGAP_Co6,
NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_Gc4, NCI_CGAP_Co6,
NCI_CGAP_Br25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with a
driver whose composition is detailed below: NCI_CGAP_Kid3
pool 1 LAM 3334-3337, 3682-3683, 3798-3803 (IMAGE
Clones 1322376-1323911, 1456008-1456775, 1500552-1502855
) NCI_CGAP_Kid5 pool 1 LAM 3338-3342, 3722-3725,
3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LAM
3575-3582, 3851-3854 (IMAGE Clones 1414920-1417991,
1520904-1522439) NCI_CGAP_Gc4 pool 1 LAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1
LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 121928-1220615)
NCI_CGAP_Co10 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 1144584-1145351) The resulting
subtracted library contained 530,000 recombinants.
Subtraction was performed as previously described [Bonaldo
& Lennon & Soares (1996)]. Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG LIB=NCI_CGAP_Le12
TAG_TISSUE=leiomysarcoma
TAG_SEQ=AAATC"
BASE COUNT      34 a      81 c      84 g      21 t
ORIGIN
alignment_scores:
Quality: 351.00      Length: 73
Ratio: 5.400      Gaps: 0
Percent Similarity: 89.041      Percent Identity: 86.301
alignment_block:
US-09-899-513-2 x AM014455/rev ..
align seg 1/1 to reverse of: AM014455 from: 1 to: 220
124 PROMETARGATGTPSerSerProArgSerSerAlaCysAlaAlaAlaAl 140
|||||
219 CCTATGGGGGGGTGTCTCGCGCGTAATCAGGTGTGCGCGTGGCGGC 170
|||||
140 aSerTtAlaAlaValProGlyProGlyArgLeuProAlaTrpProGlyAla 157
|||||
169 TTTCTACCGCGTCCCGGGCCAGGCGGATCCCGCGTGGCGCGTGCACA 120
|||||
157 YRGVAlaProArgAlaLeuProAlaProSerProGlyTTPArgAlaTrp 173
|||||
119 ACGGCGCCCGCGAGCTGCGTGCATTTGCGTGGCTGGCGGGCGCTGG 70
|||||
174 ProLeuProAlaTrpSerThAlaGlyGlnAlaArglyTTPProProPr 190
|||||
69 CCCCTCCCTGCTGAGACCGATGCCAAGCTCGGGGCTACATCCGCC 20
|||||
190 CATGTPProSerArgPro 196
|||||
19 ACGGTGGCGCTGCGGGCG 1
seq_name: gb_est1:AM138178

```

```

seq_documentation_block:
LOCUS      AM138178      189 bp      mRNA      linear      EST 29-OCT-1999
DEFINITION UT-H-BI-acv-e-11-0-UT.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2716124 3', mRNA sequence.
ACCESSION AM138178
VERSION   AM138178.1
KEYWORDS  EST.
SOURCE    human.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 189)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cga@bbs-remail.nih.gov
           Oligo-dT track not found, Not I site shown in beginning of sequence
           is likely internal to the message. cDNA library preparation: M.B.
           Soares Lab Clone distribution: NCI-CGAP clone distribution
           Information can be found through the I.M.A.G.E. Consortium/RLNL at:
           www-bio.lnl.gov/bbrp/image/image.html
           Seq primer: m13 forward
           POLYA-No.

```

FEATURES

```

Source      Location/Qualifiers
1. 189
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="IMAGE:2716124"
   /clone_lib="NCI_CGAP_Sub3"
   /lab_host="DH10B (Life Technologies)"
   /Note="Vector: pTZ19-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI. The
NCI_CGAP_Sub3 library is a subtracted library derived from
the NCI_CGAP_Sub1 library, which is a subtracted library
derived from BT. BI constitutes a mixture of 21
normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4,
NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_Cu11, NCI_CGAP_Le12,
NCI_CGAP_Br23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_Gc4, NCI_CGAP_Co6,
NCI_CGAP_Br25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 1322376-1323911, 1456008-1456775
, 1500552-1502855); NCI_CGAP_Kid5 pool 1 LAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1
LAM 3575-3582, 3851-3854 (IMAGE Clones 1414920-1417991,
1520904-1522439); NCI_CGAP_Gc4 pool 1 LAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
LAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 121928-1220615);
NCI_CGAP_Co10 pool 1 LAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 1144584-1145351). Subtraction
was performed as previously described [Bonaldo, Lennon &
Soares (1996)]. Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG LIB=NCI_CGAP_Gc4
TAG_TISSUE=germ cell
TAG_SEQ=AAATC"

```

```

BASE COUNT      28 a      66 c      69 g      26 t
ORIGIN
alignment_scores:

```

Quality: 276.00 Length: 63
Ratio: 5.208 Gaps: 1
Percent Similarity: 84.127 Percent Identity: 80.952

alignment_block:

US-09-899-513-2 x AM138178/rev ..

Align seg 1/1 to reverse of: AM138178 from: 1 to: 189

```
135 AlAcysAlaAlaAlaSerTyrAlaValProGlyProGlyArgLeuP 151
|||||
187 GGGTGTGGCGGCGAGTCGTACGGGCTCGGCCCAAGGCGACTGCC 138
|||||
151 oAlatTPProGlyAlaTyrGlyAlaProArgAlaLeuProAlaProSerP 168
|||||
137 CAATGGCGCCGTACATACGGCGCCCGCGCGCTTGCCTGCGCTTGCAC 88
|||||
168 roGlyTTPArgAlaTTPProLeuProAlaTTPSerThraAlaGlyAla 184
|||||
87 CTGGCTGGCGGCGCGCGCTTGCCTGCGCGCTTGCCTGCGCGCAAGCT 38
|||||
185 ArgGlyTTP.ProProProArgTTPProSerArgPro 196
|||||
37 CGGGCGATGACAACAACAATGGCTTCGCGCGCG 1
```

seq_name: gb_gss:CNS04N1Q

seq_documentation_block:

LOCUS CNS04N1Q 1026 bp DNA linear GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
121H07 of library G from Tetraodon nigroviridis, genomic survey
sequence.

VERSION AL298151

KEYWORDS GSS: genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

Roest-Crolius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

Unpublished
2 (bases 1 to 1026)

Roest-Crolius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

Unpublished
3 (bases 1 to 1026)

Genoscope.
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at

http://www.genoscope.cns.fr/tetraodon.

Location/Qualifiers

FEATURES

Source

1..1026

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="121H07"

/clone_lib="G"

/note="Genoscope sequence ID : CDBG121CD04LP1-end : 'T7"

BASE COUNT 121 a 367 c 348 g 184 t 6 others

alignment_scores:

Quality: 178.00 Length: 215
Ratio: 1.712 Gaps: 11
Percent Similarity: 48.372 Percent Identity: 30.233

alignment_block:

US-09-899-513-2 x CNS04N1Q ..

Align seg 1/1 to: CNS04N1Q from: 1 to: 1026

```
26 ValGlyAspLeuGlyAlaGlyGlyCysSerLeuTTPSP..... 38
|||||
174 GTGCTCCGCGCGCGCTCCGCTGCGCTGAGAGACATGACCTGCGCGCA 223
|||||
39 ...AspProAlaArgPheIleValValProAlaAlaTyrAlaLeuAla 54
|||||
224 CGGCTCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 273
|||||
54 euGlyLeuGlyLeuProAlaAsnValAlaAla...LeuAlaMetPheIle 69
|||||
274 GTGCGCGCGCGCTGCTGCGGACCTGCTGCTTCTTCTTCTGATGAGACG 323
|||||
70 ArgSerGlyArgLeuGlyAlaAlaLeuLeuLeuTyrLeuPheAsnLe 86
|||||
324 CGCGCGCGCGGAGCGGCGCGCTCCAGCATTCGCTTCTTCTTCTTCTTCA 373
|||||
86 uAlaLeuValAspGluPhePheThrLeuThrLeuGlnLeuThrLeuThr 102
|||||
374 GGCGCGCACCGACCTCCAGTTCGTGCTGAGACGCTGCTGCGCGCGCG 423
|||||
103 TyrThrLeuGlyLeuAlaArgArgPro..... 111
|||||
424 ACATGCTCTGACTTCACTGAGTCCGCTTCCGCGCGCGCATGTAGATC 473
|||||
112 .....ProAlaThrArgProGlyP 118
|||||
474 GTGCTGCGGTGACGAGTACATGATGATGATGATGATGATGATGATGAT 523
|||||
118 roProThrThrCysPro.....PromelaArgArgTTPSer.SerP 131
|||||
524 CGCATGAGCATCACCGCTACCGCGCGCTGCGCTGCGCGCTGCGCGAC 573
|||||
131 oArgSerSerAlaCysAlaAlaAlaAlaAlaSerTyrAlaValProGly 148
|||||
574 GCGCGCGCGCGCGCGGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 623
|||||
148 lYArgLeuProAlaTTPProGlyAlaTyrGlyAlaProArgAlaLeuPro 164
|||||
624 GCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 652
|||||
165 AlAPro...SerProGlyTTPArgAlaTTPPro.....LeuProAl 177
|||||
653 CGGCGCTTTCGCGCGGATGAGGCGCGCTGCGCGCGGAGCGCGCTGCT 702
|||||
177 atTPSerThrAlaGlyAlaArgGlyTTPProPro..... 189
|||||
703 CTGGGCTTCCGCGGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 752
|||||
```

seq_name: gb_est2:B1666779

seq_documentation_block:

LOCUS B1666779 758 bp mRNA linear EST 12-SEP-2001

DEFINITION 603291708P1 NRH_MGC_96 Homo sapiens CDNA clone IMAGE:5311321 5',

ACCESSION B1666779

VERSION B1666779.1 GI:15581012

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 758)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshitaki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
 Plate: LLM11788 row: 0 column: 02
 High quality sequence stop: 758.
 Location/Qualifiers
 1..758
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1b="IMAGE:531321"
 /clone_1b="NIH-MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: Bluescript (modified pluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (ctcgag); Oligo-dT primed using primer 5'-TTTTTTTCTTTTTCVN-3', size-selected for average insert size 2.3 kb and normalized to 10^5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."
 BASE COUNT 173 a 191 c 311 g 83 t
 ORIGIN
 alignment_scores:
 Quality: 164.00 Length: 184
 Ratio: 1.928 Gaps: 11
 Percent Similarity: 46.196 Percent Identity: 35.870
 alignment_block:
 US-09-899-513-2 x B166779/rev ..
 Align seg 1/1 to reverse of: B166779 from: 1 to: 758
 27 G1A5PLeuG1yValG1yG1ySerLeuTrpAspProAlaArgPh 43
 736 GGGCCACAGGGTGGCTGGGGCCAGCCCTG.....CCAGCCGCTT 696
 43 e11e.....ValValProAlaAlaAlaAlaAlaAlaLeuG1yLeuG 57
 635 TCTCAGCCCCGCGCTAGTGCCTGAGCTGGCCACTTCTCCAGCTCT 646
 57 LyleProAlaAsnValAlaAlaAlaAlaAlaAlaAlaAlaLeuValas 73
 645 GCCTC.....ATGCCAGACAGGCTCTCCAGCTCTGCTGCTGCTG 602
 74 ArgLeuG1yG1nAlaLeuLeuLeuLeuLeuPheAsnLeuAlaLeuValas 90
 601 CGGTTTGGGCTGCCCTCCTCTCTTCTCTTC..... 569
 90 pG1uPhePheTrLeuTrLeuG1nLeuTrLeuTrLeuTrLeuG1yL 107
 568CTCGGGG 562
 107 euAlaArgArgProProAlaTrArgProG1yProProThrThrcysPro 123
 561 TAGCTCGAGACCTGTGAGGGGCA.....GGCCGCGCTCAAGACTGTCTCT 518

124 ProMetArgArgTrpSerSerProArgSerAlaAlaAlaAla 140
 517 CCGGAGAGGATG.....GCCTCAGACCTGCTGCGAGCTGCGGC 480
 140 aser.....TyrAlaValProG1yProG1yArgLeuProAlaTrp 153
 479 CCAGGCGCCCTGAAAGCTAGCCCGGGCCGAGAGAGAGCTCATGGA 430
 154ProG1yAlaTrpG1yAlaProArgAlaLeuProAla..... 165
 429 ACTGTCGGGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAG 380
 166 ProSerProG1yTrpArgAlaTrpPro.....LeuProAlaTrpSerTh 180
 379 CCGTACAGCTCTTGGCCAGCTGCTCATTGCTCAGAGTTTGAGT.. 332
 180 rAlaG1yG1nAlaAlaArgG1yTrpProProArgTrpProSerArgProp 197
 331CTGCCACTCTCTTGGAGAGCCGCTCTC 304
 197 ro 197
 303 CT 302
 seq_name: gb_est2:BF312953
 seq_documentation_block:
 LOCUS BF312953 843 bp mRNA linear EST 21-NOV-2000
 DEFINITION 601896218F1 NIH-MGC_19 Homo sapiens CDNA clone IMAGE:4125676 5',
 TITLE mRNA sequence.
 ACCESSION BF312953
 VERSION BF312953.1 GI:11260861
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 843)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
 Plate: LCM1014 row: e column: 05
 High quality sequence stop: 802.
 Location/Qualifiers
 1..843
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1b="IMAGE:4125676"
 /clone_1b="NIH-MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pORF; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."
 BASE COUNT 185 a 249 c 267 g 142 t
 ORIGIN
 alignment_scores:
 Quality: 164.00 Length: 184
 Ratio: 1.928 Gaps: 11
 Percent Similarity: 46.196 Percent Identity: 35.870

Quality: 163.50 Length: 212
Ratio: 1.858 Gaps: 12
Percent Similarity: 41.509 Percent Identity: 32.075

alignment_block:
US-09-899-513-2 x BF312953/rev ..

Align seg 1/1 to reverse of: BF312953 from: 1 to: 843

```

26 ValGlyAspLeuGly.....ValGlyGlyCys..... 34
   |||||
588 GTAGAGCCGATAGTGGCTGGCTTGGCCAGCTGAGGCTGGCTCC 539
   |||||
35 .....SerLeuTrpAspAspProAlaArgPheLeu 45
   |||||
538 TACCGGAACAGCCTCGGCGCATCTGGCCAGACCTCCAGAGGACAG 489
   |||||
45 AlValProAlaAlaTyrAlaLeuAlaLeuGlyLeuGlyLeuProAlaAsn 61
   |||||
488 CAGAGCCAGGCGCACAGAGGT...GCCCTGGGGGCCAGCCCTGCACCGGT 442
   |||||
62 ValAlaAlaLeuAlaMetPheIleArgSerGlyGlyArgLeuGlyAla 78
   |||||
441 GTCACAGCCCGC.GCCGTAGTGCCTGCACAGCTGTGGCCACTTCTCCAGC 393
   |||||
78 AlLeuLeuLeuTyrLeuPheAsnLeuAlaLeuValAspLupPhePheTrL 95
   |||||
392 T..... 392
   |||||
95 eufhrLeuGlnLeuTrpLeuThr.....TyrTyr 104
   |||||
391 .....CTCCCTCAATGCGCGACAGGCTCTCCAGCTCGTGTCTGTGTC 347
   |||||
105 LeuGlyLeuAlaArgArgProProAlaThrArgPro..... 116
   |||||
346 TCGCGTTTGGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 297
   |||||
117 .....GlyProProThrThrCysProProMetArgArgTyrPS 129
   |||||
296 ACCGTGAGGAGGCGCGCCCTCAAGGCTGTCTCTCGGAGAGATG.... 251
   |||||
129 erSerProArgSerSerAlaCysAlaAlaAlaSer.....Tyr 142
   |||||
250 .....GCCTCCAGCTCTGTCGACGTCGCGGCGCCAGGCGCCCTGAGC 209
   |||||
143 AlValAlaProGlyProGlyArgLeuProAlaTrp.....ProGlyAlaTyr 157
   |||||
208 CGTAGGCGCGGCGCGGAGGAGAGCTTCATGAGACTGTCGCGGTTC 159
   |||||
157 rGlyAlaProArgAlaLeuProAla.....ProSerProGlyTyrPA 171
   |||||
158 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 109
   |||||
171 rGlyAlaTrpPro.....LeuProAlaTrpSerThrAlaGlyAlaArg 185
   |||||
108 CAGCTGTGTCATCTGTCACAGCTTGGAGT..... 77
   |||||
186 GlyTrpProProArgTrpProSerArgProPro 197
   |||||
76 .....CCTCCCATCTTGGAGAGCCGCTCTCT 47
   |||||
seq_name: gb_est2:BE259788
seq_documentation_block:
LOCUS BE259788 575 bp mRNA linear EST 26-OCT-2000
DEFINITION 601152574F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3508555 5',
mRNA sequence.
ACCESSION BE259788
VERSION BE259788.1 GI:9130544
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
1 (bases 1 to 575)
NIH-MGC <http://mgc.ncl.nih.gov/>.
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
plate: LICM190 row: c column: 20
High quality sequence stop: 571.

FEATURES
Source
Location/Qualifiers
1..575

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3508555"
/clone_id="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Library constructed by Ling Hong
in the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 128 a 155 c 231 g 61 t
ORIGIN

alignment_scores:
Quality: 161.00 Length: 172
Ratio: 1.872 Gaps: 9
Percent Similarity: 50.000 Percent Identity: 36.047

alignment_block:
US-09-899-513-2 x BE259788/rev ..

Align seg 1/1 to reverse of: BE259788 from: 1 to: 575

```

40 ProAlaArgPheIleValValProAlaAlaTyrAlaLeuAlaLeuGlyLe 56
   |||||
567 CCGGCGAGCGGCTGTCTCAAGCCCGCCCTAGTGCCTGCGAGCTGGGGGCC 518
   |||||
56 uGlyLeuProAlaAsnValAlaAlaLeuAlaMetPheIleArgSerGly 72
   |||||
517 ACCCTGCCAGCGCGGTCTCAAGCCCGCCCTAGTGCCTGCGAGCTGGTG 468
   |||||
73 GlyArgLeuGlyAlaLeuLeuLeuTyrLeuPheAsn.LeuAlaLeu 89
   |||||
467 GGCACATTTCTCCAGACTCTGCCTCAATGCGCGACAGCTCTCCACCTCC 418
   |||||
89 AlAspLupPhePheTrLLeuThrLeuGlnLeuTrpLeuThrTyrTyrTrL 105
   |||||
417 TGTCTCTGTGCTGTGGGTTGGCGCTCTCTCTCTCTCTCTCTCTCTCT 368
   |||||
106 GlyLeuAlaArgArgProProAlaThrArgProGlyProProThrTrhrcy 122
   |||||
367 GGGGTAGCTTCAGAGACTTGAGGGGCA.....GGCCCGCTCAAGGCTG 324
   |||||
122 sProProMetArgArgTyrSerSerProArgSerSerAlaAlaAla 139
   |||||
323 TCCTCCCGGAGAGATG.....GCCGCCAGCTCGTCGACGCTG 286
   |||||
139 lAlaSer.....TyrAlaValAlaProGlyProGlyArgLeuProAla 152
   |||||
285 CGGCCAGGCGCCCTGAGACCGCTAGCCCGGCGCGGAGAGAGAGCTTCA 236

```

